Hydrological Applications of Multi-source Soil Moisture Products

By

Moonhyuk Kwon BSc, MSc

A Dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Engineering

Department of Civil Engineering

April 2019

Approximately 46,500 words

ABSTRACT

Soil moisture (SM) is a key state variable in understanding the hydrologic processes, including runoff, infiltration, drought and crop growth. Thus, acquiring accurate SM information has been a priority in hydrology; however gathering such data remains a challenge for many areas of the world with respect to their spatio-temporal aspects. This challenge has contributed to the popularity of using complementary tools such as satellite retrievals and land surface models. This thesis examines multi-source SM data and further enhances their practical use in SM simulation.

First, the exponential filter method is introduced to estimate a root zone SM based on satellite surface SM. Subsequently, a cumulative distribution function matching approach is applied, not only to address the inevitable systematic biases between in-situ and satellite SM but also to determine an ideal temporal combination. The performance of each bias-correction group is validated through a cross-validation procedure. However, a major issue in using satellite SM data for practical applications is their coarse spatial resolution. Therefore, a multivariate stochastic SM estimation approach, based on a Gaussian-mixture nonstationary hidden Markov model, is introduced to spatially disaggregate the satellite SM data for multiple locations. It is revealed that the mean correlation coefficient of the proposed model is significantly greater than that of an ordinary regression model.

The second part of this thesis focuses on expanding the applicability of SM data in hydrological applications. I introduce a hybrid modelling framework by incorporating SM state variables obtained from the Tank model and multi-satellite sensors via a machine learning based regression technique. The enhanced performance of the hybrid model over a conventional model (the Tank model) is especially apparent in the simulation of low flows; this indicates that even though the overall contribution to runoff prediction is not significant, satellite SM products appear to help capture distinct features of the rainfall-runoff process. ERA-Interim SM data are then employed to identify the spatial and temporal characteristics of agricultural drought stemming from SM deficit. Furthermore, the copula-based Multivariate Standardized Drought Index is exploited to explicitly determine the interdependence and interaction between precipitation and SM deficiency.

Keywords | *Soil moisture, Remote sensing, bias correction, Spatial downscaling, Rainfall-runoff model, Support vector machine, Multivariate drought index, Clustering analysis*

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the requirements of the University's regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. The work is original, except where indicated by specific reference in the text. Any views expressed in the dissertation are those of the author.

SIGNED:…… DATE:……..……

ACKNOWLEDGEMENTS

First and foremost, I owe my sincere gratitude to my supervisor, Prof. Dawei Han not only for his academic guidance but also for his patience and considerable support during my PhD journey. Thanks to his enormous and close supervision, I could have built up my own idea under great confidence. The logical thinking, enthusiasm and confident attitude I have learnt from him will benefit me for my life. I should also thank Dr. Miguel A. Rico-Ramirez. He has always guided me to the right research direction and encouraged me to successfully finish my work.

My deepest gratitude should also go to Prof. Hyun-Han Kwon who gave me constructive advice throughout my PhD study, which made this thesis possible. Earnest thanks should go to Dr. Kue Bum Kim and Dr. Dongik Kim for their active collaboration during my PhD study.

I would also like to thank colleagues and friends in the research group since I joined, Dr. ASMM Rahman, Dr. Null Nanding, Dr. Joost Iwema, Dr. Fanny Sarrazin, Dr. Sherien Al-Azerji, Dr. Jun Zhang, Dr. Yang Song, Dr. Barnaby Alexander Dobson, Dr. W outer Knoben, Dr. Otto Chen, Mr. Sebastian Gnann, Ms. Ludovica Beltrame, Ms. Anna Lo Jacomo and Mr. Charles West for sharing their insights, knowledge, and discussions. My special thanks should go to Mr. Hwan-Myung Jang and his family. Every single moment I shared with them in Bristol is an unforgettable memory for the rest of my life. I acknowledge K-water for giving me an opportunity to do this PhD work and their financial support.

Last but not least, my utmost gratitude also goes to my parents and sister for their continuing support throughout my life. I owe my deepest appreciation to my wife Jiyoung Kim and two children (Douyoun and Douhyun) who supported me from the very first moment of my PhD till the final viva.

Thank you!

April 2019 (Moonhyuk Kwon)

LIST OF PUBLICATIONS

Journal papers:

1. **Moonhyuk Kwon** Hyun-Han Kwon, and Dawei Han. 2018. "A spatial downscaling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian-mixture nonstationary hidden Markov model." *Journal of hydrology*. DOI: 10.1016/j.jhydrol.2017.12.015.

2. **Moonhyuk Kwon** and Dawei Han. 2019. "Assessment of Remotely Sensed Soil Moisture Products and their Quality Improvement." *Journal of Hydro-environment.* DOI: 10.1016/j.jher.2019.04.002.

3. **Moonhyuk Kwon**, Hyun-Han Kwon and Dawei Han. 2019. "Spatio-temporal Drought Patterns of Multiple Drought Indices Based on Precipitation and Soil Moisture: A case study in South Korea." *International Journal of Climatology*. DOI: 10.1002/joc.6094.

4. **Moonhyuk Kwon,** Hyun-Han Kwon, and Dawei Han. 2019. "A Hybrid Approach Combining Conceptual Hydrological Model, Support Vector Machine and Remote Sensing Data for Rainfall-Runoff Modeling." *Water Resources Research.* (under review).

Conferences:

5. **Moonhyuk Kwon** and Dawei Han. "Validation of active and passive microwave soil moisture observations with ground-based networks in South Korea." *Weather Radar and Hydrology 2017*, Seoul, Korea (Poster presentation).

6. **Moonhyuk Kwon**, Hyun-Han Kwon, and Dawei Han. "A Spatial Downscaling of Soil Moisture from Rainfall, Temperature, and AMSR2 Using a Gaussian-Mixture Nonstationary Hidden Markov Model." *EGU General Assembly 2018*, Vienna, Austria. (Poster presentation).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
AMSR2	Advanced Microwave Scanning Radiometer 2
ANN	Artificial Neural Networks
ASCAT	Advanced Scatterometer
BIC	Bayesian Information Criterion
CDF	Cumulative Distribution Function
CV	Coefficient of Variation
ECMWF	European Centre for Medium-Range Weather Forecasts
EM	Expectation-maximization
ESA	European Space Agency
FDR	Frequency Domain Reflectometry
GLDAS	Global Land Data Assimilation System
GM-NHMM	Gaussian Mixture Nonstationary Hidden Markov Model
ISMN	International Soil Moisture Network
JAXA	Japan Aerospace Exploration Agency
KMA	Korea Meteorological Administration
K-water	Korea Water Resources Corporation
LSSVM	List Squares Support Vector Machine
MODIS	Moderate Resolution Imaging Spectro-radiometer
MSDI	Multivariate Standardized Drought Index
NDVI	Normalized Difference Vegetation Index
NSE	Nash–Sutcliffe Efficiency
PACF	Partial Autocorrelation Function
PDF	Probability Density Function

PDSI	Palmer Drought Severity Index
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RR	Rainfall-Runoff
RZSM	Root Zone Soil Moisture
SM	Soil Moisture
SMAP	Soil Moisture Active Passive
SMMR	Scanning Multichannel Microwave Radiometer
SMOS	Soil Moisture Ocean Salinity Satellite
SPI	Standardized Precipitation Index
SSI	Standardized Soil Moisture Index
SSM/I	Special Sensor Microwave Imager
SWI	Soil Water Index
TDR	Time Domain Reflectometry
TIR	Thermal Infrared
TRMM	Tropical Rainfall Measuring Mission
ubRMSE	Unbiased Root Mean Squared Error
WMO	World Meteorological Organization

CHAPTER 1 Introduction

1.1 Background and motivation

Soil moisture (SM) is a key state variable for understanding hydrologic processes, including runoff, infiltration, drought, crop growth, and many other phenomena closely related to soil conditions (Albergel et al., 2008a; Barrett and Petropoulos, 2013; Brocca et al., 2011; Zhao and Li, 2013). Thus, acquiring accurate SM information has been a priority in hydrology, meteorology, and climatology. SM data can be obtained from ground-based measurement, satellite observation and SM accounting model, as well as an integration of different sources of data to address each method's limitation. In-situ observation is generally recognized as a tool for gaining the most reliable SM information and commonly used as a reference variable for hydrological applications (Dorigo et al., 2011). However, in many parts of the world, it remains challenging to collect spatially and temporally suitable ground-based SM data (Brocca et al., 2017; Peng et al., 2017; Zhuo and Han, 2016). Another practical issue in using in-situ observation is that hydrological analysis is typically implemented on a catchment scale, while point-based measurements tend to be poorly representative of the spatial distribution for a large-scale estimation of SM, due to heterogeneous land surface (Griesfeller et al., 2016; Merlin et al., 2012; Reichle et al., 2007).

Alternatively, remotely sensed SM has become a valuable complementary tool for monitoring SM conditions, providing the advantage of large-scale and relatively high temporal coverage (Brocca et al., 2011; Zeng et al., 2015). Despite the potential benefits of satellite-based remote sensing techniques, they are only able to monitor a very thin soil layer, while the root zone SM (RZSM) provides more meaningful information for hydrological applications, such as drought monitoring, crop-yielding prediction and rainfall-runoff (RR) modelling (Ford et al., 2014). Another major challenge in using satellite SM data for practical applications is their coarse spatial resolution and uncertainties stemming from an inability to resolve sub-grid scale variability. Considering the spatial mismatch and different measurement depths, a cumulative distribution function (CDF) matching method, as well as an exponential filter method, are employed in this thesis to improve the applicability of satellite-derived SM. In terms of relatively low spatial resolution of satellite SM products, various statistical approaches have used a downscaling framework to achieve a

higher spatial resolution for microwave SM data (Merlin et al., 2012; Peng et al., 2016; Piles et al., 2014; Ranney et al., 2015; Zhao and Li, 2013). However, the existing approaches all largely depend on a linear or nonlinear regression model to spatially downscale the satellite SM products, without considering the uncertain factors affecting SM dynamics. Moreover, the spatio-temporal dynamics of SM content result from complicated and mutually related processes of hydro-meteorological elements, such as subsurface flow, lateral flow, infiltration, precipitation, climate, and soil (Botter et al., 2007; Ridolfi et al., 2003). In this regard, it can be beneficial to use a multivariate Gaussian mixture nonstationary hidden Markov model (GM-NHMM) with the intention of providing a practical tool for the estimation of daily SM at the catchment scale.

Given that SM governs the interaction between rainfall and the corresponding hydrological processes, SM data obtained from various approaches can be used for enhancing hydrological applications including runoff modelling and drought monitoring. To date, satellite SM retrievals have been integrated into conceptual (Brocca et al., 2012; Massari et al., 2015) or physically-based RR models (Lievens et al., 2015; Loizu et al., 2018; Young et al., 2017) using data assimilation techniques. However, the results from these approaches vary widely, demonstrating either no (or limited) improvement in the RR simulating performance or significant improvement. In this thesis, satellite SM data are blended into a hydrological RR model by introducing a hybrid scheme that updates state variables obtained from a conceptual model (Tank) and satellite SM data via the least squares support vector machine (LSSVM) technique. Moreover, considering that a deficit of moisture in the soil is closely related to an agricultural drought (Dracup et al., 1980) and a drought is a multidimensional phenomenon, combining multiple variables (e.g., precipitation and SM) is beneficial for successful drought preparedness and mitigation, and particularly effective for communication purposes between different types of drought. In this thesis, I introduce the copula-based Multivariate Standardized Drought Index (MSDI), which combines meteorological and agricultural droughts, and the blended drought index is then grouped by the hierarchical agglomerative clustering approach for classifying regional patterns. As for drought assessment, SM data have been rarely applied, due to a lack of observation sites in many parts of the world including South Korea, and this thesis provides a useful guideline for detecting the nature of drought, especially for data-sparse regions.

Given this background, this thesis aims to enhance the understanding of multi-source SM products and further improve their applicability in hydrological application especially for RR modelling and drought assessment.

1.2 Scope and aim of this study

The goal of this PhD study is to explore multi-source SM data and further identify their potentials and limitations in the context of hydrological applications. This thesis mainly focuses on the effectiveness of the utilization of SM data in hydro-meteorological issues, which still remains at the early stage of development in many parts of the world including South Korea. The main objectives of this thesis are listed as follows:

(1) Evaluate microwave-based SM products and identify how their performance in South Korea is different from other parts of the world.

(2) Explore the pros and cons of individual SM products with the intention of selecting the optimal datasets among various sources for different hydrological applications.

(3) Investigate the feasibility of disaggregating satellite SM data through a stochastic downscaling modelling framework.

(4) Develop a methodology to incorporate SM information into a hydrological model and examine in which aspects the proposed model outperforms the conventional model.

(5) Explore the benefit of combining multiple variables (i.e., precipitation and SM) to better understand the nature of drought, hypothesizing that drought is a multidimensional phenomenon.

1.3 Thesis layout and structure

The main objective of this thesis is to enhance the applicability of SM obtained from multiple sources for the hydrological application. The structure of this thesis and its associated contents are presented in Figure 1-1. A literature review related to this study is carried out in Chapter 2. Chapter 3 describes the study area and datasets such as in-situ, satellite and ERA-Interim SM data. The main parts can be partitioned into two subparts. The first part (Chapter 4-5) is related to the assessment of satellite data and their quality improvement for practical uses in hydrology. The

second part includes Chapter 6-7 in which SM data are used for extended hydrological applications (i.e., hydrological RR modelling and drought analysis). Brief outlines of each chapter are provided below.

Chapter 2 undertakes a brief review of various SM measurement methods providing the basic knowledge of each method as well as their potential advantages and limitations in the context of hydrological applications. This chapter introduces bias correction methods that are widely used in the field of remote sensing, and it presents a brief introduction of hydrological models and drought assessments for further extended practical applications of SM information.

In Chapter 3, the study areas and datasets used in this thesis are presented. This study is primarily based in South Korea. However, this thesis focuses on either national scale or catchment scale, depending on the study purposes and/or data availability.

Chapter 4 aims to assess remotely sensed SM products and suggest a way to improve their quality for practical purposes. To do this, I first estimate the accuracy of the original satellite SM retrievals in terms of their orbits as well as variation patterns. Afterwards, the exponential filter, eventually combined with the CDF matching method, is employed to derive the RZSM, which appears to be more meaningful than the surface SM for hydrological applications.

Chapter 5 introduces a multivariate stochastic SM estimation approach based on a GM-NHMM to spatially disaggregate the Advanced Microwave Scanning Radiometer 2 (AMSR2) SM data for multiple locations in the Yongdam catchment. Rainfall and air temperature are considered as additional predictors in the proposed modelling framework.

Chapter 6 and Chapter 7 aim to expand the applicability of SM data for practical purposes in the context of the hydrological application. In this thesis, I primarily focus on two emerging applications that are currently receiving growing attention in hydrology.

Chapter 6 introduces a hybrid RR modelling tool that updates SM state variables obtained from a conceptual model (Tank) by combining the LSSVM model within a calibration process. The fundamental idea behind this approach is that different models could capture different aspects of the RR process.

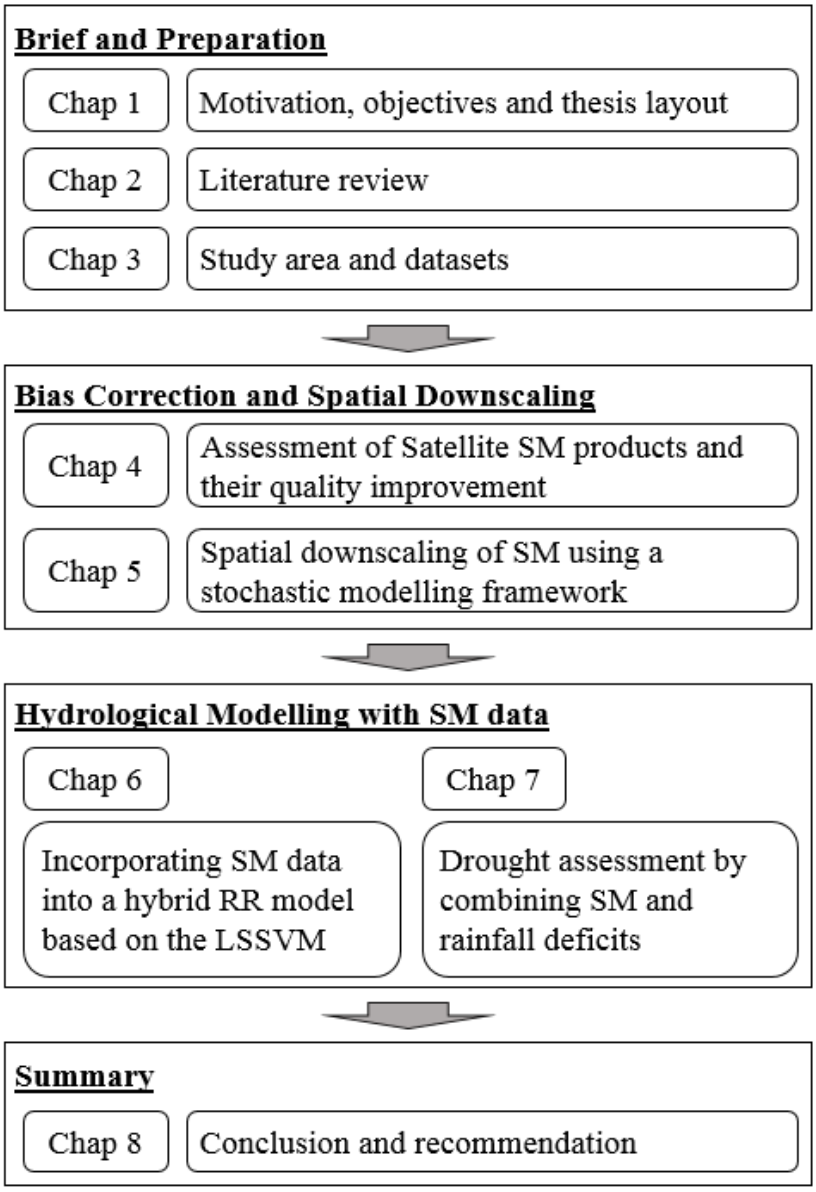


Figure 1-1 Layout and structure.

In Chapter 7, historical droughts are assessed by the copula-based MSDI that combines the Standardized Precipitation Index (SPI) and Standardized Soil Moisture Index (SSI) in the context of a multivariate probability distribution, by considering the interdependence and interaction between rainfall and SM. Further, the relationships between the SPI and SSI indices as well as the transition from meteorological to agricultural droughts are explored over the past three decades (1986-2016) in South Korea.

The conclusions and recommendations based on the major findings of this thesis are presented in Chapter 8, followed by a discussion of future work.

CHAPTER 2 Literature review

2.1 Introduction

As mentioned in Chapter 1, this thesis mainly focuses on exploring multi-source SM data and improving their applicability for a better understanding of hydrological processes that are particularly associated with SM information. This chapter introduces the basic knowledge of SM measuring methods and reviews on pre-processing schemes of SM data together with their emerging hydrological applications. To be specific, this chapter is organised as follows. Section 2.2 provides the fundamental background of SM measurement methods ranging from in-situ observations to their alternative data sources such as satellite- and model-based SM. The concept of pre-processing schemes that are required to address the inherent limitations of satellite SM is given in Section 2.3. This section also introduces popular downscaling schemes and their pros and cons in the context of practical applications. Section 2.4 reviews the literature on RR modelling, beginning with a conceptual model, and then the recent advances in the data-driven approach, especially for the use of machine learning techniques. The role of SM in identifying historical droughts and current research trends in assessing the characteristics of droughts are presented in Section 2.4.

2.2 Soil moisture measuring methods

SM estimates can be obtained in multiple ways including point measurements, remote sensing techniques, and land surface models as well as an integration of different data sources to address the limitations of the individual approach. Site-specific observations of SM are commonly regarded as the most accurate estimates but gathering such data is generally expensive and time-consuming (Peng et al., 2017). In addition, one notable shortcoming of using point-based measurements is that they are rarely representative of the spatial distribution for a large-scale catchment. Therefore, in-situ observations are often used for calibrating and testing hydrologic models. Satellite SM data, on the other hand, cover wider areas with relatively high temporal resolution, providing SM information for the top few centimetres of the soil layer.

A notable technique in measuring SM content is the use of the cosmic-ray neutron sensor that bridges the gap in spatial resolution between conventional point observations and remote sensing-based SM measurements. Since the first construction of SM networks in the United States by the University of Arizona, the cosmic-ray neutron technique has gained popularity in measuring field-scale RZSM in the hydrological community, providing continuous SM information over a wide range of land covers, soil types and climate conditions (Brunetti et al., 2019; Nguyen et al., 2019). Up until now, virtually 200 stationary cosmic-ray neutron detector systems are in operation over the world and this innovative approach has been proven a useful tool for practical applications such as hydrological modelling, data assimilation and reference datasets when assessing satellite SM products (Andreasen et al., 2017). Despite the abovementioned potential, the direct use of the cosmic-ray neutron sensor still remains a challenge since the accuracy in measuring SM contents is easily contaminated by the presence of surrounding hydrogen sources such as vegetation and snow (Andreasen et al., 2017). Another alternative way to monitor SM conditions is to integrate different source of data based on a data assimilation scheme. For instance, the Global Land Data Assimilation System (GLDAS) and ERA-Interim SM data are globally available with high spatio-temporal resolution. More details on various SM data sources are presented in the following sections.

2.2.1 In-situ observations

In-situ observations can be collected through either direct or indirect methods. In terms of a direct approach (also known as the Thermo-gravimetric technique), a sample of soil is taken for lab analysis, and then the water in the soil sample is removed, by either evaporation or a chemical process (Petropoulos et al., 2014). In this manner, SM content is directly measured by the difference in weight between before and after the removal of water. This classical SM measurement technique is considered inexpensive and straightforward but can be destructive since it would not be feasible to repeatedly measure SM at the same spot (S.U et al., 2014). Therefore, this method is widely used for calibrating other SM measurement sensors instead of being employed for the construction of SM networks. In other words, SM contents obtained from this method is commonly regarded as a standard reference (S.U et al., 2014). Table 2-1 shows a summary of in-situ SM observation techniques and more details on each approach is presented as follows.

Table 2-1 Summary of in-situ SM observation techniques. This table is presented based on S.U et al. (2014).

Sensor	Flexibility	principle	Unit	Remarks
Thermo-gravimetric	Lab scale	Evaporation/ chemical action	Gravimetric	High accuracy, no health risk, time-consuming, destructive test
Neutron	In-situ	Neutron scattering	Volumetric	Health risk, immediate response, more suitable in subsurface soil
Dielectric (TDR)	In-situ/lab	Dielectric constant	Volumetric	No health risk, immediate response, easy to use, fail in highly saline soils, relatively independent soil type

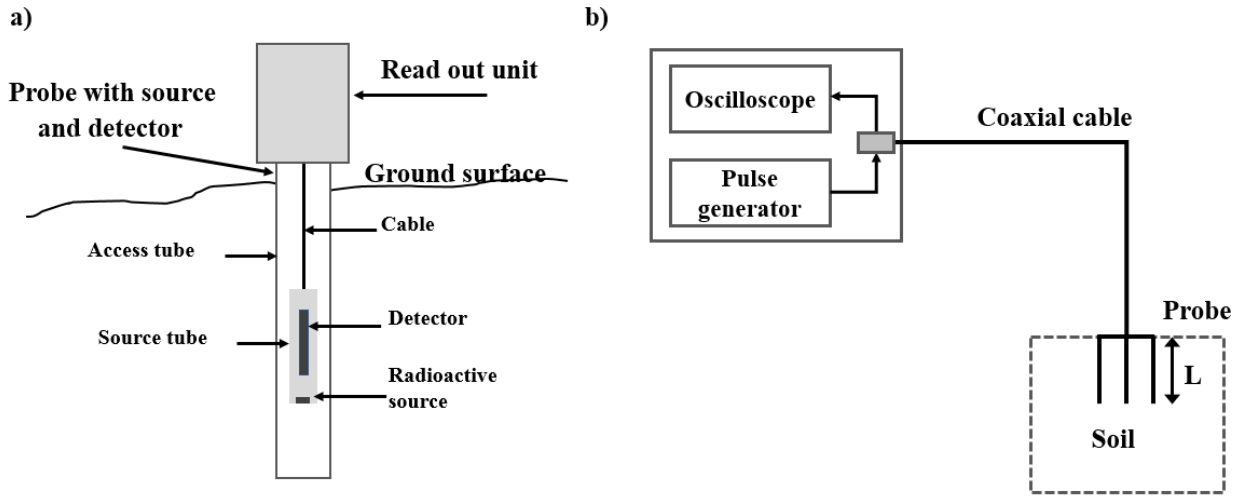


Figure 2-1 Schematic diagram depicting SM measurements using a) the neutron probe and b) the TDR (adapted from Su et al., 2014).

2.2.1.1 Neutron scattering technique

The neutron scattering technique is an indirect and non-destructive SM measurement scheme (Bell, 1968). This technique uses a neutron probe that detects thermal neutrons for measuring SM conditions, providing both surface and profile volumetric SM content. The neutron SM measurement sensor is made up of a probe and a gauge component (Figure 2-1a). A source of fast neutrons is emitted by the probe, and the gauge detects the flux of slowed down neutron energy

that has hit hydrogen. The detection rate is dependent on the amount of water in the soil since the energy of the emitted fast neutrons are significantly slowed down during the process of collisions between the neutrons and the hydrogen in the soil (S.U et al., 2014). This approach basically measures the amount of hydrogen content in the soil and it can be regarded as estimating moisture content since the presence of hydrogen in the soil is an indication of water contents (Bell, 1968). One of the advantages of using this method is the ability to measure at several depths offering the characteristics of the SM profile, while relatively high installation costs, low spatial resolution and health hazards associated radiation exposure need to be addressed (Jarvis and Leeds‐Harrison, 1987).

2.2.1.2 Dielectric techniques

SM measurement technique that uses a dielectric property of the soil involves Time Domain Reflectometry (TDR) capacitance technique and Frequency Domain Reflectometry (FDR). Given that the basic concept of the two techniques is the same, I briefly introduce the fundamental theory of the TDR method. Over the last decades, the TDR has been widely used for the measurement of SM content since this approach is capable of continuously providing in-situ measurements at the same location in a non-destructive manner with a high level of reliability. The fact that the accuracy of this approach is rarely affected by temperature makes it more useful to measure SM in shallow soils (S.U et al., 2014).

This technique is based on the concept that there exists a considerable gap in dielectric constant between dry soils and pure water. Thus, changes in soil dielectric permittivity measured by the TDR sensor are directly related to changes in SM contents in the soil. As can be seen in Figure 2-1b, the pulse generator sends a voltage pulse to a probe installed in the soil along a coaxial line and the TDR measures the velocity of an electromagnetic wave. The velocity varies considerably depending on the dielectric constant of the soil, which is primarily sensitive to the water content in the soil. The dielectric constant k_a is calculated based on Eq. 2-1 (Topp et al., 1980).

$$K_a = \left[c\frac{t}{2L}\right]^2 \quad (2\text{-}1)$$

where c is the velocity of the light ($3 \times 10^8\ m/s$) and t is the travel time of an electromagnetic pulse through the probe length (L). Then, the volumetric SM content is determined using an

empirical formula. For instance, Topp (1980) proposed an empirical polynomial equation, formulating the relationship between K_a and SM content (θ) as follows:

$$\theta = -5.3 \times 10^{2} + 2.29 \times 10^{-2} K_a - 5.5 \times 10^{-4} (k_a)^2 + 4.3 \times 10^{-6} (k_a)^3 \quad (2\text{-}2)$$

This equation is known to be applicable for various soil conditions, but it is still required to build a site-specific relationship between SM content and dielectric constant by a calibration process for different regions (S.U et al., 2014).

2.2.2 Remote sensing

Figure 2-2 Global distribution of SM stations available at the ISMN from January 2000 (accessed on 24 May 2018).

Up until now, significant efforts have been made to construct a long-term SM observation network. For instance, the International Soil Moisture Network (ISMN) created a web-based data collection platform by serving in-situ measurements from individual SM observation networks over the world (http://www.ipf.tuwien.ac.at/insitu). The ISMN provides approximately 6,100 SM datasets from more than 1,400 observation sites obtained from globally-scattered 35 SM networks (Dorigo et al., 2013). Although in-situ observations have increasingly become available, gathering such data over large areas is still challenging in many parts of the world (Figure 2-2). The issues such

as the lack of SM observations and relatively short-term observation periods have, in turn, led to gaining the popularity of using remotely sensed SM products.

Table 2-2 Summary of remote sensing techniques. This table is presented based on Wang and Qu (2009) and Moran et al. (2004).

Group (wavelength)		Property observed	Advantage	Shortcomings
Optical (0.4 – 2.5 μm)		Soil reflection	Fine spatial resolution Broad coverage	Sensitive to soil type Cloud contamination
Thermal infrared (3.5 – 14 μm)		Earth Surface temperature	High spatial resolution Physically well described	Bare soil only Very top soil layer only Cloud contamination Contaminated by surface topography
Microwave (1 – 30 cm)	Passive	Brightness temperature	All weather Penetrates some vegetation Broad coverage Better SM sensitivity Physically well understood	Coarse spatial resolution Contaminated by surface roughness and vegetation Interference from communications
	Active	Backscatter coefficient Dielectric constant	All weather Relatively high spatial resolution Physically well described	Low temporal resolution limited swath width Contaminated by surface roughness and vegetation

A remote sensing technique for measuring SM can be categorised into three groups (i.e., optical, thermal and microwave) in terms of the electromagnetic spectrum (Sadeghi et al., 1984). Over the last few decades, remote sensing technology has been becoming an important complementary tool for observing and monitoring hydrological components including SM retrievals, especially where observation data are scarce or not available. Table 2-2 presents a summary of the pros and cons of remote sensing techniques for the measurement of SM contents. More details are described in the following subsections.

2.2.2.1 Optical and thermal infrared

Optical remote sensing technique uses a strong relationship between reflected radiation from the surface in the visual spectrum and the amount of water in the soil (Sadeghi et al., 1984). This approach is designed to build an empirical relationship between them. However, this empirical approach tends to be location-dependent, resulting in a wide disparity when applied to other regions that have different soil conditions compared to the calibration regions. This is due to the

fact that the spectral characteristic of soil is sensitive to other factors (e.g., mineral composition, organic matter, soil texture, and surface roughness) (Petropoulos et al., 2015). It is also found that unlike microwave sensor that uses longer wavelength, the optical signal is easily contaminated by cloud and vegetation cover, making this approach less prevalent for SM estimation (Petropoulos et al., 2015).

Thermal infrared remote sensing measures the heat emission (i.e., Earth radiative temperature) of the surface, which is eventually used for estimating SM content typically by combining vegetation indices (e.g., the normalised vegetation index) (Wang and Qu, 2009). The fundamental principle behind this approach is that land surface temperature is significantly related to the variation of surface SM content, i.e., in areas with high SM content, the surface temperature is warmer at night, while it tends to be cooler during the day (Van de Griend and Engman, 1985). This method has the advantages of higher spatial resolution, while the inherent drawback of this approach is that Earth radiative temperature is considerably affected not only by SM content but also by many factors such as vegetation, wind speed, and humidity (S.U et al., 2014).

2.2.2.2 Microwave

Microwave-based SM retrieval techniques utilise the considerable differences in dielectric properties between dry soil ($k_a \approx 4$) and water ($k_a \approx 80$), i.e., the effective dielectric constant significantly rises as the amount of water in the soil increases, thus showing high sensitivity to SM content (Engman and Gurney, 1991; Wagner et al., 2013). During the last several decades, microwave-based active and passive sensors have been used to retrieve SM content by measuring backscattering and emission from the Earth surface, respectively. Compared with other remote sensing techniques that use visible and infrared radiation, microwave remote sensing techniques using longer wavelengths have a more significant potential to measure SM in that they are mostly unaffected by weather conditions such as cloud cover, haze, rainfall, and aerosols (Barrett and Petropoulos, 2013; Chauhan et al., 2003). However, it has been reported that the accuracy of remotely sensed SM is affected by many factors such as soil property, surface roughness and vegetation coverage (Schmugge, 1978). For instance, Wang et al. (2016) compared satellite SM with normalised difference vegetation index (NDVI) as proxy of vegetation development. They found that the correlation coefficients between satellite SM and NDVI were poor in the higher vegetation density region. Dorigo et al. (2012) obtained better performance in measuring SM

content over sparsely vegetated areas. Based on the comparison with ground-based SM networks and land surface model, recent studies have also shown that more reliable satellite-based SM retrievals can be obtained in areas with less vegetation impact (Champagne et al., 2010; Hain et al., 2011).

Historically, the Scanning Multichannel Microwave Radiometer (SMMR) on board the Nimbus-7 satellite provided microwave-based geophysical variables including SM observation at a global scale for the first time, which was operated between 1978 and 1987 (Reichle et al., 2007). The Special Sensor Microwave Imager (SSM/I) was then launched in July 1987 followed by the Tropical Rainfall Measuring Mission (TRMM) that has been operating since its launch in 1997. Several space missions employing microwave remote sensing are presently in operation, offering near real-time surface SM measurements (Brocca et al., 2017) (Table 2-3). The European Space Agency's (ESA) Soil Moisture Ocean Salinity Satellite (SMOS) launched in 2009 is the first SM dedicated satellite, mainly focusing on measuring surface SM (Kerr et al., 2012). Compared to abovementioned satellites, the SMOS detects the brightness temperature using a longer wavelength region (L-band; 1.4 GHz), which makes it able to observe SM contents at a deeper depth (approximately 5 cm of soil) (Ford et al., 2014). Most recently, NASA has begun providing SM measurements at a global scale with high spatio-temporal resolution (2-3 day revisit frequency) using both L-band active and L-band passive microwave (Entekhabi et al., 2010).

There are also two other sensors that have been widely used for SM retrievals from space: the AMSR2 on board the Global Change Observation Mission (GCOM)-W1 satellite (JAXA, 2013) and the Advanced Scatterometer (ASCAT) on board the Meteorological Operational (METOP) satellite (Albergel et al., 2008b). In the preliminary study, I attempted to evaluate SMOS and SMAP SM products. However, the number of SM data acquired from both satellites was too small for their effective evaluation. Based on practical considerations (i.e., data availability), this thesis is dedicated to evaluating satellite SM products from the ASCAT and AMSR2. In the past few decades, many studies have been conducted to examine the accuracy of active and passive microwave sensors and to expand their applicability for practical issues in hydrology. For example, Wu et al. (2016) evaluated the AMSR2 by analysing ascending and descending overpass products to each other as well as comparing 598 in-situ SM observation stations from the ISMN. Their findings have revealed that AMSR2 SM retrievals tend to underestimate in-situ measurements,

and similar results were obtained by Zeng et al. (2015) over the Tibetan Plateau region. In contrast to the AMSR2, which uses passive microwave sensing techniques, the ASCAT provides a global satellite-based active microwave SM product. Validation studies based on the ASCAT have been mainly carried out across Europe, and the results show that the ASCAT could produce SM with a reasonable level of accuracy (Albergel et al., 2008a; Brocca et al., 2010; Wagner et al., 2013; among others). However, the performance of satellite-based SM products varies from region to region since their accuracy in measuring SM contents is severely contaminated by many factors. In this regard, the accuracy and reliability are assessed in Chapter 4 prior to their practical applications.

Table 2-3 Summary of presently operational satellite missions.

	ASCAT (Active)	**AMSR2 (Passive)**	**SMOS (Passive)**	**SMAP (Active & Passive)**
Frequency	C-band	Multiple bands	L-band	L-band
Spatial resolution	12.5 km, 25 km	10 km, 25 km	35-50 km	3-40 km
Temporal resolution	3 days	1-2 days	1-3 days	3 days
Observation depth	0.5-2 cm	~ 1 cm	~ 5 cm	~ 5 cm
Mission period	Since 2007	Since 2012	Since 2009	Since 2015
Reference	Albergel et al., 2008b	JAXA, 2013	Kerr et al., 2012	Das et al., 2011

2.2.3 Model-based (reanalysis) SM products

Currently, several modelling frameworks continuously produce globally operational SM datasets with relatively high spatial-temporal resolution, which includes the GLDAS, ERA-Interim and the Modern-Era Retrospective Analysis for Research and Applications. These SM products are typically derived on the basis of assimilation techniques by combining both in-situ measurements and modelling outputs (Paiva et al., 2016).

The GLDAS has been developed to provide land surface states (e.g., Evapotranspiration, snow depth, SM and soil temperature) and their fluxes using data assimilation techniques (https://ldas.gsfc.nasa.gov/index.php). The GLDAS is composed of four different land surface models; the Mosaic model, the Noah model, the Community Land Model and the Variable Infiltration Capacity model (Rui and Beaudoing, 2017). The four land surface models in the

GLDAS platform produces SM datasets at different soil layers. For instance, the Noah model provides SM information at four soil layers (0–10 cm, 10–40 cm, 40–100 cm, and 100–200 cm), while SM contents in three soil layers (0–2 cm, 2–150 cm, and 150–350 cm) can be obtained from the Mosaic model. For more details, readers are kindly referred to Rui and Beaudoing (2017). The GLDAS SM products can be received from the GLDAS data server (https://disc.gsfc.nasa.gov/datasets). The meteorological parameters are produced by assimilating multiple data sources including ground-based measurements, satellite observations and model-based products. The GLDAS provides global SM datasets daily in a 3-h time interval with 0.25˚ to 1.0 ˚ spatial resolution. The 1.0° data are available from 1979 onwards, while the 0.25° data can be obtained since 2000.

The European Centre for Medium-Range Weather Forecasts (ECMWF) has produced the ERA-Interim SM data on the basis of a sequential data assimilation technique (Dee et al., 2011). In this approach, in-situ measurements are also incorporated into the assimilation framework, and this approach not only provides SM information but also produces a variety of hydrological variables such as snow depth, temperature and soil temperature (Dee et al., 2011). The ECMWF releases global reanalysis SM datasets daily in quasi-real time with high spatial resolution, in 6-hour intervals (Albergel et al., 2012). Unlike microwave sensors that only provide SM information of a shallow near surface soil layer, the ERA-Interim offers SM datasets at four depths (i.e., 0–7, 7–28, 28–100 and 100–289 cm), covering from the surface to the RZSM. Thus, SM information obtained from the ERA-Interim can be more applicable for RZSM-based studies, especially in the field of meteorological modelling and hydrology. ERA-Interim SM data with a spatial resolution of 25 km covering the period 1979 – present can be accessed from https://www.ecmwf.int/. The accuracy of the ERA-Interim reanalysis data was assessed against in-situ observations from 117 stations across the world by Albergel et al. (2012). Their results revealed robustness for various climate conditions with a high level of accuracy. Zeng *et al.* (2015) compared the ERA-Interim SM data against multiple remotely sensed SM products (i.e., the AMSR2, SMOS and ASCAT) and they found the ERA-Interim SM showed the best agreement with in-situ observations.

2.3 Pre-processing approaches

2.3.1 Scaling technique

The mismatch in spatial resolution and observation depths between in-situ observations and satellite retrievals leads to inevitable systematic biases. Thus, prior to the actual use of remotely sensed SM, a statistical bias correction approach is regarded as a crucial step for the subsequent use (Kornelsen and Coulibaly, 2015). In order to match the variability of SM obtained from the space with in-situ observations, various methods have been proposed. Here, I briefly introduce bias correction methods that have been widely used in the field of remote sensing: the CDF matching, linear scaling and Min/Max correction method.

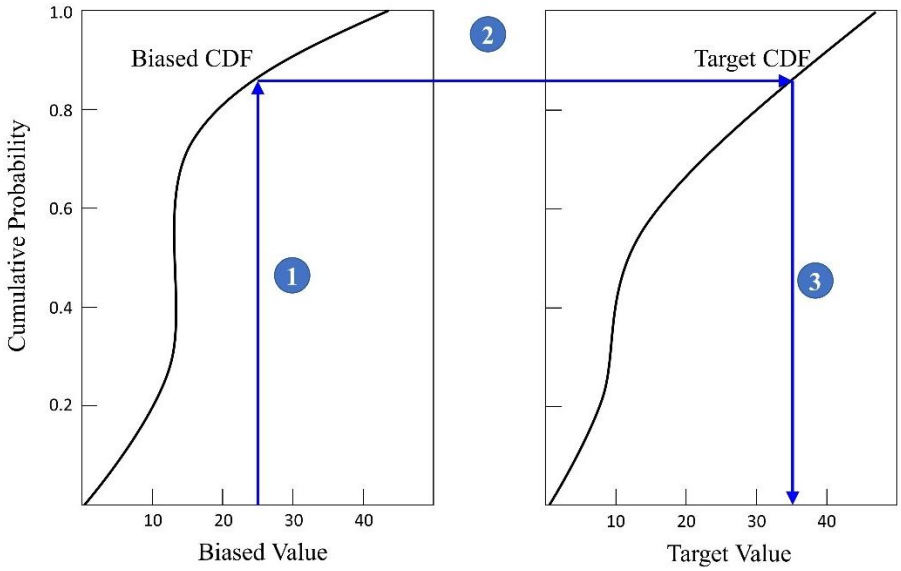


Figure 2-3 Schematic representation of the CDF matching procedure.

One common approach is the CDF matching method that is considered to be an enhanced nonlinear technique applied to tackle systematic differences between two datasets (Su et al., 2013; Brocca et al., 2011; Liu et al., 2011; Scipal et al., 2008). This bias correction method is also commonly adopted in the field of climate change studies. Through this method, the CDF of the biased data is rescaled and eventually matched with that of the target measurements (Figure 2-3). In other words, the biased products are mapped to the same probability value as that of observations (target dataset). The CDF matching method can be expressed as follows:

$$SM_{cdf} = F_{obs}^{-1}\left(F_{sat}\{SM_{sat}(t)\}\right) \quad (2\text{-}3)$$

where SM_{sat} is the biased data (satellite retrieval), SM_{cdf} is the bias corrected data (CDF matched value), F_{sat} is a CDF of biased data, and F_{obs}^{-1} is an objective CDF (the CDF of the observed).

The linear scaling approach is used to adjust the statistical characteristics of biased data. Namely, the mean (μ) and standard deviation (σ) of biased data are rescaled to match those of in-situ observations (SM_{obs}).

$$SM_{reg} = \frac{[SM_{sat} - \mu(SM_{sat})]}{\sigma(SM_{sat})}\sigma(SM_{obs}) + \mu(SM_{obs}) \quad (2\text{-}4)$$

The third approach involves the Min/Max correction method. Here, the maximum and minimum values of biased values are rescaled with intent to match those of the reference time series.

$$SM_{minmax} = \frac{[SM_{sat} - \min(SM_{sat})]}{[\max(SM_{sat}) - \min(SM_{sat})]}[\max(SM_{obs}) - \min(SM_{obs})] + \min(SM_{obs}) \quad (2\text{-}5)$$

The fundamental assumption behind these bias correction methods is that the agreement of temporal dynamics between in-situ and satellite data is more important than the absolute retrieval values (Kornelsen and Coulibaly, 2015). In this thesis, the scaling (bias correction) technique is not only used for validation of satellite SM (Chapter 4) but also employed for a pre-processing procedure (Chapter 5).

2.3.2 Filtering technique

Satellite SM products based on the C- (for ASCAT) and X- band (for AMSR2) are representative of only very top soil layer, while ground-based measurements are generally measured at deeper layers. This issue associated with the disparity in observation depths can be effectively addressed by employing filtering techniques, i.e., this technique allows one to drive the RZSM from surface SM obtained from remote sensing.

Many attempts have been made to estimate the RZSM from SM at a top layer. For instance, data assimilation techniques, such as Extended Kalman Filter and Ensemble Kalman Filters, were introduced to combine satellite surface SM with a different source of data to reproduce the RZSM (Renzullo et al., 2014; Sabater et al., 2007). Additionally, Zaman and Mckee (2014) used a machine learning scheme to estimate the RZSM by assimilating surface SM, soil temperature and

precipitation datasets. However, above-mentioned schemes have a high computational cost (González-Zamora et al., 2016). Alternatively, the exponential filter method used in this thesis, also known as Soil Water Index (SWI), proposed by Wagner et al. (1999), has been widely used owing to its simplicity of implementation, computational efficiency and robustness for representing the RZSM (Albergel et al., 2008a, 2008b; Ceballos et al., 2005; Ford et al., 2014; Paulik et al., 2014; Qiu et al., 2014). This scheme assumes that a soil profile consists of the surface layer and subsurface layer, and the SM dynamics of the lower layer is proportionally linked with the difference between the two layers. A recursive formulation of the exponential filter that is relatively straightforward to implement but provides a mathematically equivalent principle to the original filter method is adopted in this study following Albergel et al. (2008b):

$$SWI_n = SWI_{n-1} + K_n \left[SSM_{(t_n)} - SWI_{n-1}\right] \tag{2-6}$$

where SWI_n is the estimated profile SM. t_n and t_{n-1} are the observation times of the current and the previous SSM measurement in Julian days. The filtering is initialized by applying SWI_0 = $SSM_{(t_0)}$ and K_0 = 1, respectively. $SSM_{(t_n)}$ refers to the surface SM estimate at t_n, and the gain K at time t_n is calculated as follows:

$$K_n = \frac{K_{n-1}}{K_{n-1} + e^{-\frac{(t_n - t_{n-1})}{T}}} \tag{2-7}$$

where T is a surrogate parameter (also known as characteristic time) that characterizes the temporal dynamics of SM along the soil profile. The performance of this approach is highly dependent on the T parameter that normally increases with the depth of observation. In other words, the time series of the SWI becomes smoother since a larger T value is generally expected to be representative of SM contents at a deeper soil layer. This scheme is also used in hydrology for a pre-processing step when satellite-derived SM products are assimilated into RR models (Brocca et al., 2012; Loizu et al., 2018; Massari et al., 2014).

The exponential filter method is applied when satellite SM data are compared against in-situ observations (Chapter 4). Additionally, this technique is used as a pre-processing task prior to incorporating satellite SM data into a hybrid RR model in Chapter 6.

2.3.3 Spatial downscaling

As stated in Section 2.2, the satellite remote sensing technique has been employed to measure surface SM, and numerous efforts have been made to improve the reliability and accuracy of satellite SM retrievals. Yet, one major challenge in using satellite SM data for regional applications in hydrology is their coarse spatial resolution and uncertainties stemming from an inability to resolve sub-grid scale variability. To overcome those limitations, various statistical approaches have used a downscaling scheme to achieve a higher spatial resolution for microwave SM data (Merlin et al., 2012; Peng et al., 2016; Piles et al., 2014; Ranney et al., 2015; Zhao and Li, 2013).

The downscaling methods can be categorised depending on the type of data being studied, such as radar, optical/thermal, topography, or soil information data (Peng et al., 2017). Optical/thermal sensor data (generally vegetation index, surface temperature, albedo, etc.) have been widely used to disaggregate the original satellite SM products into fine-scale estimates because they not only provide land surface parameters at higher spatial resolution (Peng et al., 2016; Piles et al., 2011; Zhao and Li, 2013) but also have a significant correlation with SM (Fang and Lakshmi, 2014; Peng et al., 2015; Srivastava et al., 2013). The basic idea behind these approaches is to build a statistical model (based on the relationship between the satellite SM products and surface parameters) that can simulate SM sequences using given surface parameters as predictors. As mentioned in Section 2.2, the most frequently reported limitation of this approach is that optical and thermal properties can be obtained only under clear-sky conditions (Djamai et al., 2016; Park et al., 2017). Geo-information data, such as topography, soil attributes, and vegetation, have also been used to disaggregate coarse-scale SM values into fine-scale ones using a regression framework (Busch et al., 2012; Ranney et al., 2015). Given that this approach empirically establishes the relationships between coarse-scale SM products and fine-scale geo-information data, the established downscaling framework appears to be location-dependent, which hampers its applicability for a broader range of applications (Peng et al., 2017).

Recently, machine learning techniques have been used to spatially downscale satellite-based SM data for enhanced spatial resolution (Im et al., 2016; Park et al., 2017; Srivastava et al., 2013; Xing et al., 2017). For example, Srivastava et al. (2013) tested and compared several machine learning techniques, including an artificial neural network, a support vector machine, and a relevance vector machine, to spatially downscale SMOS SM datasets. Specifically, they used Moderate Resolution

Imaging Spectro-radiometer (MODIS) land surface temperature as auxiliary information in disaggregating the SMOS SM products. Park et al. (2017) developed a downscaling scheme based on a modified regression tree model that combined multiple sensors (i.e., the AMSR2 and ASCAT) with four other predictors: MODIS land surface temperature, the normalised difference vegetation index, land cover, and a digital elevation model.

Most of the existing approaches have attempted to exploit a linear or nonlinear regression modelling tool to spatially downscale microwave-based SM products without considering the uncertain factors affecting SM dynamics. Moreover, the influence of spatio-temporal variability in precipitation and temperature has been rarely considered within the downscaling framework (no literature regarding SM has been found), which could better represent the behaviour of catchment-scale SM. In this regard, this thesis introduces a stochastic downscaling approach to spatially disaggregate the AMSR2 SM retrievals. In this approach, both air temperature and precipitation are additionally considered as predictors (Chapter 5).

2.4 Emerging hydrological applications using soil moisture

Given an essential role of SM in the hydrological cycle, SM data has been widely exploited for various study purposes such as numerical weather prediction (Scipal et al., 2008), rainfall estimation (Schneider et al., 2014), RR prediction (Crow and Ryu, 2008; Massari et al., 2018) and drought monitoring (Rahmani et al., 2016). This thesis primarily concentrates on the latter two hydrological applications among others.

2.4.1 Rainfall-runoff model

A hydrologic model is an abstract representation of a complex hydrologic system that facilitates understanding and predicting water resources in a simplified manner. Notably, the RR model plays a central role in different aspects of water management, such as streamflow record extension for designing hydraulic structures (Curran, 2012), streamflow prediction for ungauged catchments (McIntyre et al., 2005) and assessment of the impact of land use or climate change (Jakeman, 1993; Pfannerstill et al., 2014; Shrestha, 2009). Over the last three decades, the development and evaluation of a runoff prediction model have attracted much attention in the field of hydrology. Although considerable efforts and improvements of RR models have been made to achieve a better

understanding of catchment response to rainfall events, there still remains the question whether these improvements meet scientific and practical demands (Orth et al., 2015). A spatio-temporal variability in the RR process appears to be significantly spatially dependence, i.e., the runoff response to rainfall is intricately linked to various influencing factors such as climate condition, land cover, soil property, topography and even human activities including irrigation and urbanization (Devia et al., 2015; Jakeman, 1993; Kisi and Parmar, 2016).

RR models can be in general categorized into three main groups: data-driven, conceptual and physically-based model (Devia et al., 2015; Shrestha, 2009). Data-driven models, also known as an empirical or black-box model, directly build RR relationships with historical records, particularly through the use of statistical models and machine learning schemes on relatively large samples. In other words, data-driven models can be used to represent the RR relationships governing catchment response to rainfall, in a statistical sense at various levels of abstraction and hence can be used to predict runoff sequences based on a set of predictors over time (Behzad et al., 2009; Young et al., 2017) . In contrast to data-driven models, conceptual models generally consist of a series of interconnected systems that represent hydrological processes at different spatial and temporal scales in an abstract manner (Paik et al., 2005; Song et al., 2017), while physically-based models are formulated in terms of mathematical (or physical) relations and interactions among hydrologic processes to transform rainfall into runoff (Beven and Kirkby, 1979; Young et al., 2017).

In this section, I review the literature on RR modelling, beginning with the conceptual model, covering work in the 1970s-2000s, and then the recent advances in data-driven approaches, mainly through the use of machine learning techniques. Conceptual models, such as Sacramento model (Burnash et al., 1973), IHACRES (Jakeman, 1993), PDM model (Moore, 1985) and HBV model (Bergström, 1976) typically require fewer parameters compared to physically-based models (e.g., MIKE-SHE (Butts and Graham, 2005), SWAT (Neitsch et al., 2011), and TOPMODEL (Beven and Kirkby, 1979)). This leads to less computational complexity and provides near real time predictions. Conceptual models are therefore relatively straightforward to construct and implement while providing comparable estimates of runoff at the lower computational cost compared to physically-based models (Orth et al., 2015; Vaze et al., 2011). A comparative study of conceptual models was carried out by the WMO during the period 1967-1974. This project compared 10

operational hydrological models and provided not only a comparison of the result of streamflow estimations but also a comprehensive description of the model structures (Sittner, 1976).

According to Orth et al. (2015), the degree of complexity and sophistication of the hydrological model does not ensure an improvement in performance, which is in turn more likely to rely on either hydrological variables (e.g., rainfall, runoff and temperature) or initial (or antecedent) wetness of the catchment. Among them, the Tank model proposed by Sugawara (1979) has gained popularity in the hydrological community primarily due to the simplicity of its structure and reasonable accuracy, even in the absence of SM information. In this respect, this well-tested model is found to be effective and useful for both practical and theoretical aspects of RR modelling (Basri, 2013; Fumikazu et al., 2013; Paik et al., 2005; Song et al., 2017). To be specific, the Tank model has been officially adopted for simulating streamflow on a daily basis over the whole of South Korea to support long-term water resources management and planning every 5-10 years. Therefore, the Tank model was selected for subsequent experiments in this thesis (Chapter 6).

Over the past few decades, machine learning techniques (e.g., SVM, fuzzy logic model, artificial neural networks (ANN) and genetic algorithm) have been widely used for hydrological modelling. Among these methods, SVM has been applied for solving classification problems (Samui and Kothari, 2011) in the field of hydrology and for predicting runoff (Bray and Han, 2004; Granata et al., 2016; Wu et al., 2014, among others). A notable trend in the context of data-driven modelling is to combine machine learning schemes with the existing conceptual/physically-based models, providing a new platform for RR modelling by taking advantage of the individual approach. For instance, Sivapragasam et al. (2001) combined the singular spectrum analysis with the SVM that provides improved accuracy in RR modelling compared to the sole use of a nonlinear regression model. Young et al. (2017) proposed a hybrid model by jointly modelling the Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS) and two different types of machine learning techniques (i.e. ANN and SVM). In a similar context, some approaches that combine the existing RR models with machine learning techniques can be found in the literature (e.g., Fernando et al., 2012; Hosseini and Mahjouri, 2016).

Given that SM governs the interactions between rainfall and the corresponding hydrological processes, such as evapotranspiration, infiltration, and percolation, the accuracy in RR modelling tends to depend on how the RR model simulates spatially and temporally accurate SM conditions

in the study area (Brocca et al., 2010; Dorigo et al., 2017; Loizu et al., 2018; Massari et al., 2015). However, insufficient in-situ observations prevent their use in the hydrological modelling framework in many parts of the world. Alternatively, assimilating satellite SM data into hydrological models have gained popularity in the hydrological community, and many studies can be found in the literature (Lievens et al., 2015; Loizu et al., 2018; Young et al., 2017; among others). The results from these studies vary widely, demonstrating either no (or limited) improvement in RR modelling to significant improvement. Given the above-mentioned background, this thesis introduces a hybrid RR modelling framework that incorporates SM state variables obtained from both a satellite and a conceptual model (i.e. the Tank) via a machine learning based regression technique in Chapter 6.

2.4.2 Drought assessment

Drought is a periodic phenomenon that exerts multifaceted negative impacts on a wide range of water-related sectors, which can eventually lead to severe direct (or indirect) socio-economic losses (Mckee et al., 1993; Spinoni et al., 2014; Vidal et al., 2009). Droughts occur virtually everywhere in the world, but their characteristics, such as duration, intensity and frequency, vary significantly depending on climate zones (Mirabbasi et al., 2013). Additionally, it is expected that climate change will accelerate changes in drought characteristics (Dai, 2011; Van Loon et al., 2016). Thus, drought monitoring and early warning systems at global and local scales have emerged as powerful platforms for preventing and mitigating the effects of droughts.

Many studies have been conducted to estimate the onset, persistence and termination of drought events using meteorological and hydrological variables (Ganguli and Ganguly, 2016; Mo, 2011; Shukla et al., 2011; among others). Drought features such as duration, severity and intensity are commonly characterized by drought indices and the degree of drought is assessed by the predefined threshold, which was first introduced by Yevjevich (1967). Drought index provides a more comprehensive perspective for drought monitoring and management compared to the direct use of hydro-meteorological variables (e.g., precipitation, SM and streamflow) (Zargar et al., 2011). Nonetheless, the selection of the drought index for a certain purpose remains controversial (Farahmand and AghaKouchak, 2015). Specifically, the identification of drought can be attributed to the choice of drought index that, with some limitations, incorporates different aspects of drought conditions (Hao and Singh, 2015). Accordingly, various drought indices have been proposed to

detect different types of droughts. For example, a meteorological drought index refers to deficits in precipitation and/or evaporation, whereas agricultural and hydrological drought indices are based on deficiencies in SM and streamflow, respectively (Dracup et al., 1980).

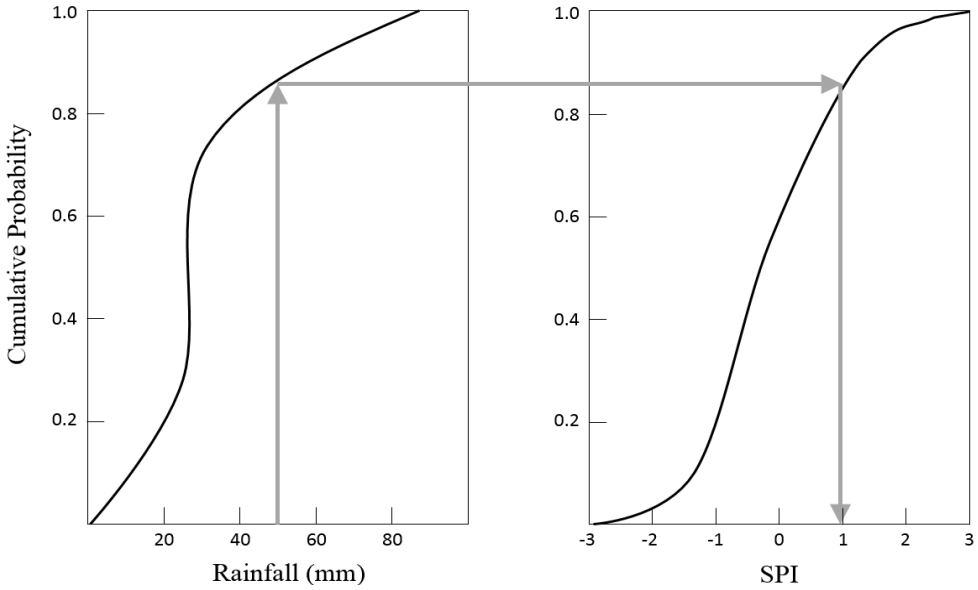


Figure 2-4 Schematic representation determining the SPI.

Many efforts have been made to develop a more effective drought index to better characterize spatio-temporal drought patterns (Hao and AghaKouchak, 2013; Karavitis et al., 2011; Vidal et al., 2009). The drought indices are mainly used to describe different types of droughts (i.e., meteorological, agricultural, hydrological, and socioeconomic drought). They are commonly derived from using a single hydrological variable (e.g., rainfall: the standardized precipitation index and standardized anomaly index, streamflow: standardized streamflow index and streamflow drought index, groundwater: standardized water-level index and soil moisture: Standardized soil moisture index). On the other hand, there are several examples that combine two or more variables such as the Palmer drought severity index (PDSI), standardized precipitation evapotranspiration index (SPEI), surface water supply index (SWSI) and multivariate standardized drought index (MSDI). Here, I briefly present the summary of several drought indices in Table 2-4, for more details readers are kindly referred to Svoboda and Fuchs (2017).

Among many drought indices, the SPI, proposed by Mckee et al., (1993), is one of the most widely used indices for the evaluation of drought conditions in the practical aspects of drought monitoring.

More specifically, the advantage of the SPI method lies in its relative simplicity of computation and ease of interpretation, and the SPI approach is particularly useful in predicting drought onset (Hao and AghaKouchak, 2013). Thus, the SPI was endorsed as a standard drought indicator by the World Meteorological Organization (WMO) (Hayes et al., 2011). The fundamental idea behind the SPI can be employed for other hydrometeorological variables with the objective of building a Standardized drought index (Kumar et al., 2016; Van Loon, 2015). Figure 2-4 schematically presents the methodology to derive the SPI: long-term rainfall data are first fitted to theoretical probability distribution functions. Then, the SPI is computed by transforming the cumulative probability distribution into a standardized normal distribution with zero mean and standard deviation equal to one (Guttman, 1999; Mckee et al., 1993). This probabilistic approach is based on Guerrero-Salazar and Jevjevich (1975) and has drawn significant attention in the field of drought assessment.

Table 2-4 Summary of drought indices. This table is presented based on Svoboda and Fuchs, (2017).

Indices	Input parameters	Strengths	Weaknesses
Standardized Precipitation Index (SPI)	P	Can be calculated over multiple timescales	Uses precipitation data only, i.e., it only detects the rainfall deficit.
Standardized Soil Moisture Index (SSI)	SM	Can be calculated over multiple timescales	Uses SM data only
Palmer Drought Severity Index (PDSI)	P, T, AC	Considers both water supply (precipitation) and demand (potential evapotranspiration)	Does not present accurate results in mountainous, frozen and snow covered areas.
Standardized Precipitation Evapotranspiration Index (SPEI)	P, T	Accounts for the influence of temperature changes	Requires a large amount of complete dataset for temperature and precipitation
Palmer Hydrological Drought Severity Index (PHDI)	P, T, AC	Allows comprehensive water system to be considered	Does not consider human impacts
Surface Water Supply Index (SWSI)	P, RV, SF, SP	Considers snow pack streamflow, rainfall and reservoir storage	Basin-dependent (cannot be compared with different regions)
Multivariate Standardized Drought Index (MSDI)	MI	Can be used to monitor more than just drought.	Requires serially multiple complete dataset

Note: P: precipitation, SM: soil moisture, T: temperature, AC: available water content, RV: reservoir storage, SF: streamflow, SP: snowpack, MI: multiple indicators.

Each drought index has limitations and strengths in measuring drought conditions. For instance, previous studies have revealed that the SPI is more likely to detect the emergence of drought conditions, whereas drought persistence can be more effectively identified based on the SM deficit. (Mo, 2011). In other words, drought information based on a single index may not be adequate for providing an integrated picture of different types of drought (Kao and Govindaraju, 2010). This issue has, in turn, activated the need for a combination of multiple drought indices that are derived from different hydrologic variables including rainfall, SM, groundwater and streamflow. Such hydrological variables particularly have been used to construct a joint drought index to characterize the complex nature of drought (Hao and Singh, 2015; Mirabbasi et al., 2012). Over the past decade, many researchers have proposed statistical models to build a multivariate drought index (Farahmand and AghaKouchak, 2015; Hao and AghaKouchak, 2013; Kao and Govindaraju, 2010; Mirabbasi et al., 2013). In the context of multivariate analysis, copulas have been used in a wide range of hydrological applications such as multivariate drought frequency analysis (Ekanayake and Perera, 2014; Kao and Govindaraju, 2010; Kwon et al., 2016; Mirabbasi et al., 2012; Sprenger et al., 2016), flood frequency/risk analysis (Favre et al., 2004; Jongman et al., 2014; Zhang and Singh, 2006), and rainfall simulation (Li et al., 2013). In Chapter 7, because of the interdependence and interaction between rainfall and SM, the copula-based MSDI is employed to consider the two drought indices (i.e., the SPI and SSI) jointly in the context of a multivariate probability distribution.

As for the large-scale evaluation of drought, it is necessary to clarify regional trends in drought rather than point-based assessment, which conveys a valuable knowledge to public stakeholders or decision makers. A number of previous studies have been carried out to investigate regional drought patterns using cluster analysis, principal component analysis (PCA) or a combination of the two techniques (Bloomfield et al., 2015). Among them, clustering analysis, also known as an unsupervised classification model, is widely used to classify drought patterns into particular categories according to their relationships; such analysis can identify specific spatio-temporal patterns within the cluster (Shamshirband et al., 2015). Santos et al. (2010) employed K-means clustering and PCA to explore spatial and temporal patterns of the SPI series, whereas Yoo et al. (2012) applied the K-means approach to partitioning their study region into several sub-regions based on bivariate drought attributes. Furthermore, spatio-temporal drought patterns were regionally summarised by combining a quantile regression model and the hierarchical

agglomerative clustering algorithm (Shiau and Lin, 2016; Yang et al., 2017). Despite the above-mentioned potential uncertainty in identifying drought features using a drought index based on a single variable, clustering analysis has not been applied extensively for a multivariate drought index. In other words, most studies on clustering analysis were dedicated to the delineation of homogeneous regions using a single drought index or their drought characteristics (e.g., duration, severity, and frequency). Moreover, given that the nonstationarity in drought episodes is of increasing concern, the nonparametric Mann-Kendall (MK) test has been widely used to identify significant changes in drought pattern (Güner Bacanli, 2017; Subash and Ram Mohan, 2011). Despite its popularity in the hydrological community, the MK approach cannot be applied to explore the temporal variability of hydrologic variables at various quantiles of the distribution (Shiau and Lin, 2016), which is essential for water resources management, especially for extreme rainfall that translates into both droughts and floods. In this regard, the quantile regression model proposed by Koenker and Bassett (1978) is introduced in Chapter 7, which aims to explore the non-Gaussian distribution of trend in drought characteristics in terms of the predefined quantiles (e.g., moderate, severe and extreme drought).

CHAPTER 3 Study Sites and Datasets

3.1 Study sites

This thesis has been conducted based in South Korea. However, two different study sites, ranging from national to catchment scale, are used, depending on the study's purpose and/or data availability.

3.1.1 Description of South Korea: national scale

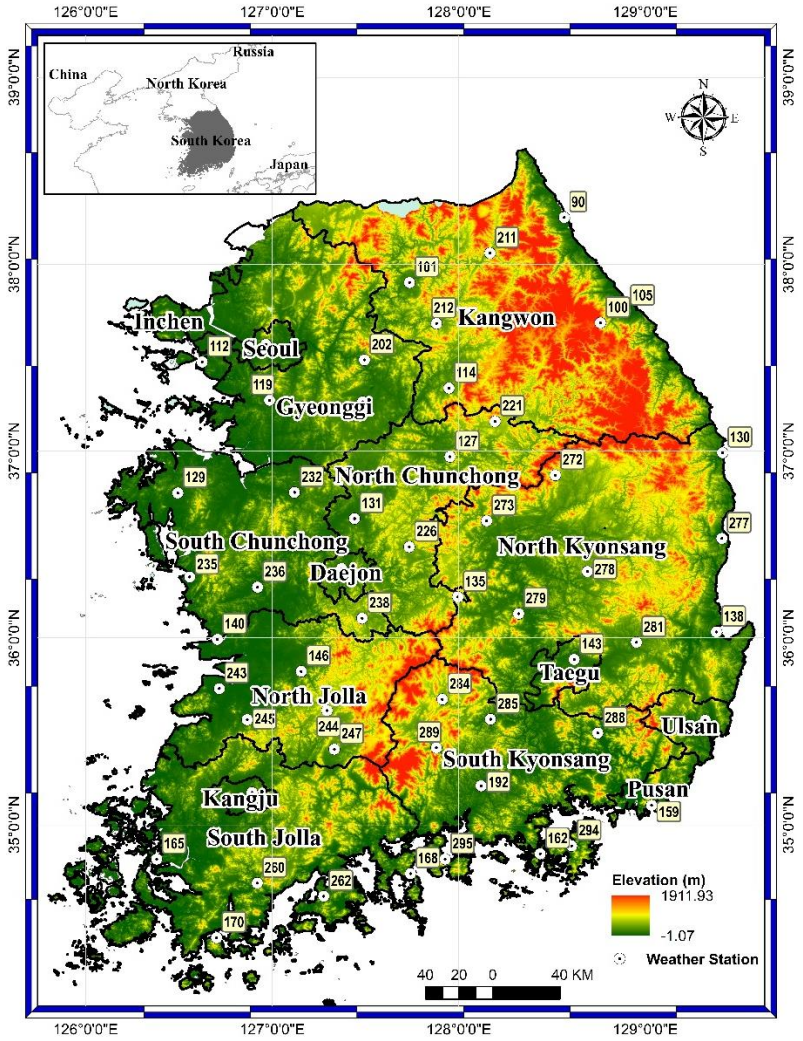


Figure 3-1 Topographic map showing the locations of the provinces in South Korea, together with weather stations.

South Korea, located in northeast Asia, has a range of 33°-38°N latitude and 124°-131°E longitude and is divided into nine provinces and seven metropolitan cities. Figure 3-1 shows the location of the provinces in South Korea along with 55 weather stations throughout the country. South Korea's climate is characterised by a cold, relatively dry winter and a hot, humid summer, displaying the typical characteristic of monsoons. As shown in Figure 3-2a, two-thirds of the annual rainfall (1,277 mm) comes during the flood season (between June and September), and only one-fifth of the rainfall comes during the dry season (from November to April), leading to challenging conditions for effective water resources management. Typhoons accompanied by extreme rainfall cause severe economic losses and environmental damage in the summer and early autumn, and South Korea often suffers from droughts before the summer monsoon begins. Note that Figure 3-2 and 3-3 are presented based on the *National Water Resources Long-term Plan,* established by the Ministry of Land, Infrastructure and Transport (MOLIT) of the Korean government in 2016.

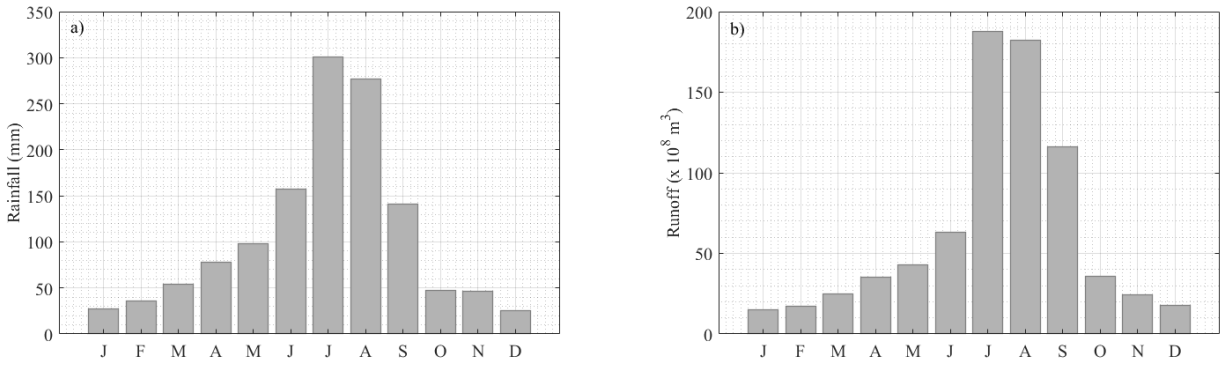


Figure 3-2 a) Monthly rainfall and b) Monthly Discharge in South Korea (MOLIT, 2016).

As shown in Figure 3-3, the total renewable water resources are about 132.3 billion m^3, while actual renewable water accounts for 57 % (76.0 billion m^3). Due to the intensive characteristics of rainfall, 41 % of the total renewable water resources are flood runoff, while runoff during the non-flood season accounts for 16 % (21.2 billion m^3). Approximately 29 % of the annual runoff goes to the sea, accounting for 38.8 billion m^3. Total yearly water demand is 37.2 billion m^3, of which 20.9 billion m^3 is supplied by dams scattered across South Korea, having a key role in water supply. The country has a total area of 99,720 km^2, of which about 64 % (63,721 km^2) is mountainous, especially along the east coast. The plains are mainly located in the west and south coasts, accounting for 18.1% of the total area (https://www.cia.gov/library/publications/the-world-factbook/geos/ks.html).

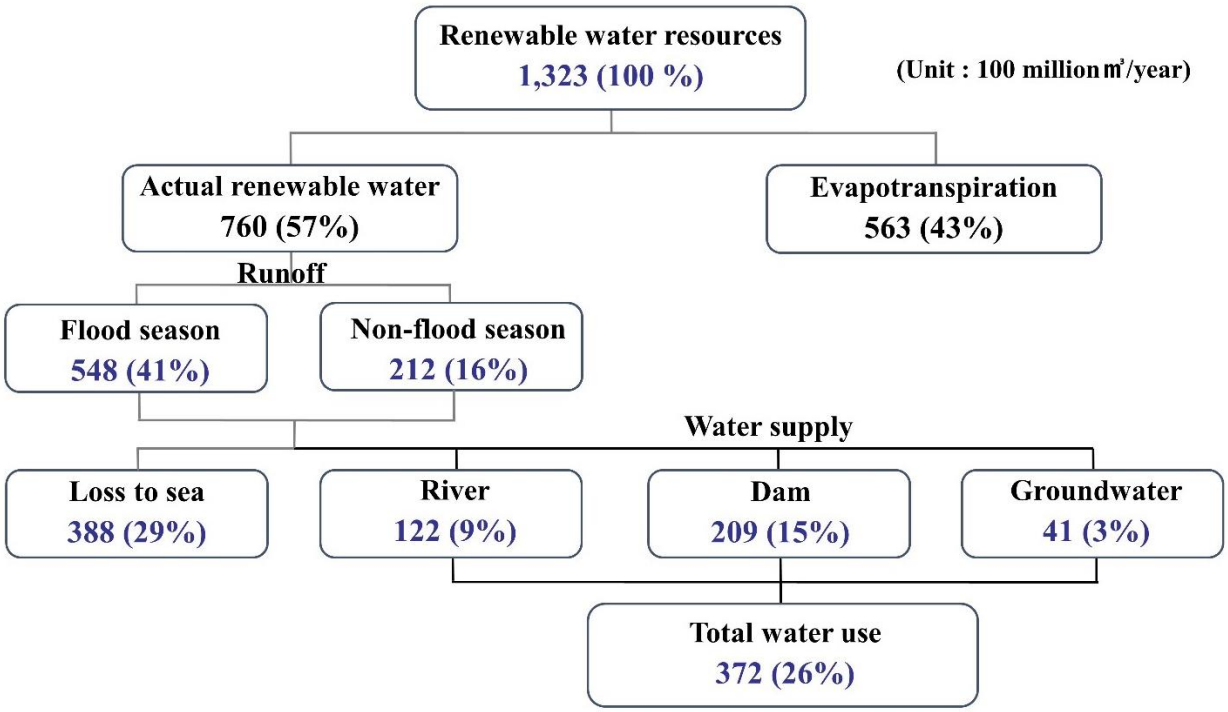


Figure 3-3 Renewable water resources and water use in South Korea (MOLIT, 2016).

3.1.2 Description of the Yongdam catchment: catchment scale

In addition to national scale, some of the studies in this thesis (Chapter 5 for downscaling and Chapter 6 for RR prediction) are carried out based on the Yongdam catchment (35.6°–36.0°N latitude and 127.3°–127.7°E longitude). Figure 3-4 shows the study area and six in-situ SM stations where precipitation data are also measured by Korea Water Resources Corporation (K-water; http://www.ydew. or.kr/kdrum/main/main.do). Additionally, air temperature (available for download from https://data.kma.go.kr/cmmn/main.do) is measured at the Jangsu weather station operated by the Korea Meteorological Administration (KMA; https://web.kma.go.kr/eng/). The SM observation network covers a drainage area of 930 km^2 with elevation ranging from 209 to 1,588 m a.s.l. Figure 3-5a presents the map of land cover, showing that the majority of the area is covered by forest (70.1%), and the rest is covered by urban (2.6 %), agriculture (19.2 %) and grass (3.3 %). The dominant soil types consist of sand (62.1 %), loam (20.7%), and silt (17.0 %), as shown in Figure 3-5b. The characteristics of land use and soil texture distribution of the Youngdam catchment is summarised in Table 3-1.

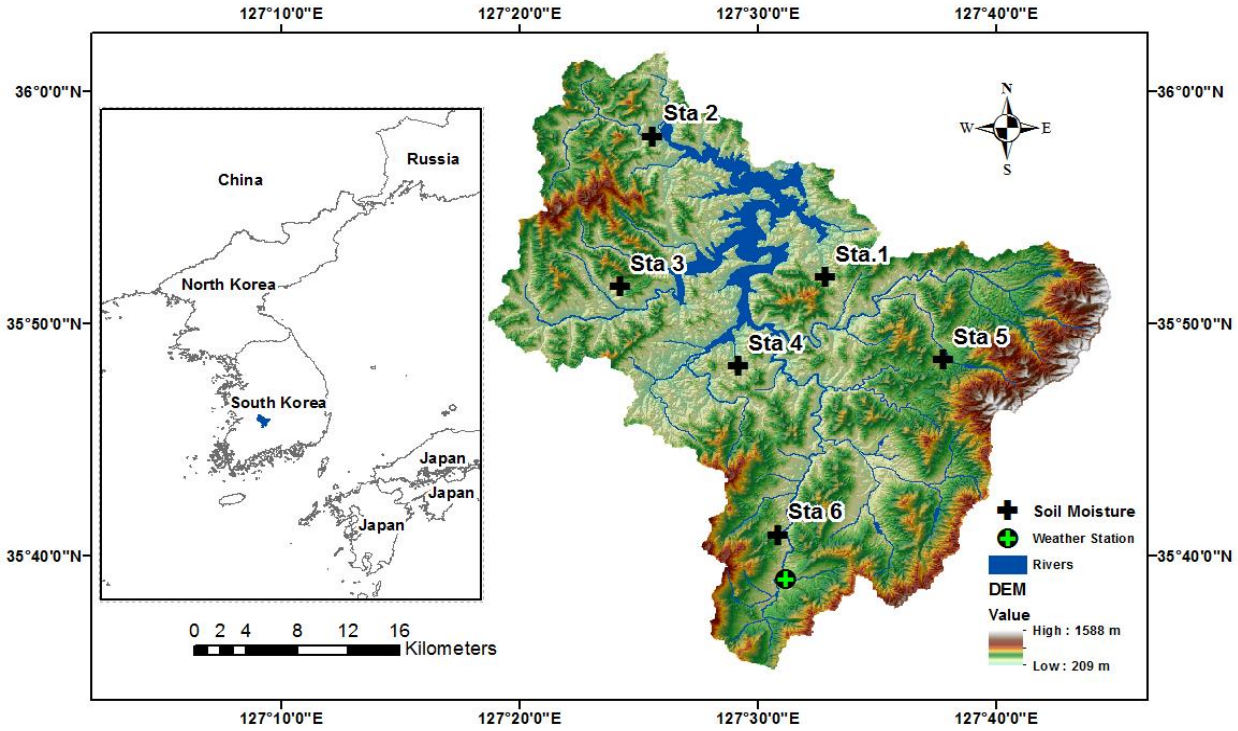


Figure 3-4 Yongdam catchment with topography and observation stations.

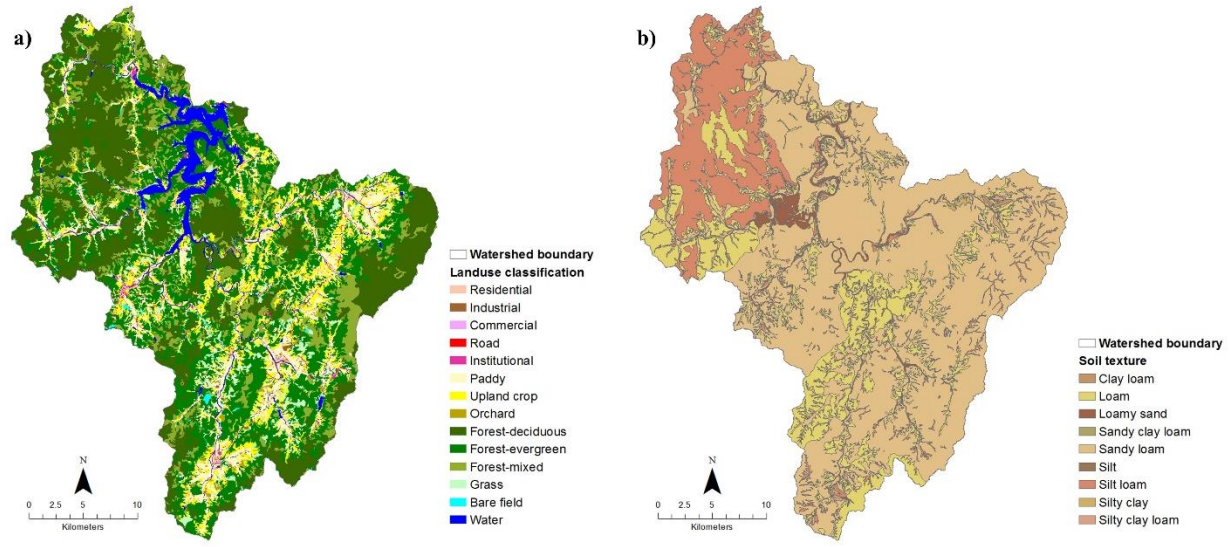


Figure 3-5 Map showing a) land use and b) soil texture in the Yongdam catchment (source: K-water).

The Yongdam dam was built for multi purposes (water supply (1,050,000 m^3/day), power generations (24,400 kW) and flood control) at the outlet of the catchment and has been operated by K-water (http://english.kwater.or.kr) since its construction in 2000 (Yoo, 2009). The pondage of the dam is 815 million m^3, with a height of 70 m and a length of 498 m (Figure 3-6).

Figure 3-6 Image of the Yongdam dam (source: K-water).

Table 3-1 Land use and soil texture distribution of the Youngdam catchment.

Land use	**Urban**	**Agriculture**	**Forest**	**Grass**	**others**
km^2 (%)	23.6 (2.6 %)	177.1 (19.2 %)	646.4 (70.1%)	30.1 (3.3)	52.8 (4.8%)
Soil type	**Clay**	**Loam**	**Sand**	**Silt**	
km^2 (%)	0.93 (0.1%)	192.5 (20.7%)	577.5 (62.1%)	158.1 (17.0%)	

3.2 Datasets

3.2.1 Rainfall data

This thesis uses rainfall data obtained from two organisations in South Korea: 1) the KMA (for national scale) and 2) the K-water (for catchment scale). Rainfall data obtained from the K-water are employed for catchment scale approaches (Chapter 5 and Chapter 6), while rainfall records measured by the KMA are used for national scale analysis (Chapter 7). To be specific, in Chapter 5, historical rainfall records from the K-water during the investigation period (2014-2016) are used to spatially disaggregate the AMSR2 SM in the Youngdam catchment. The same source of rainfall

data, but with a more extended period (2007-2016) are used as inputs to explore catchment response to rainfall in Chapter 6. As for the KMA data, the historical daily rainfall data measured at 55 weather stations, are collected for the period 1986-2016, and these are then adopted to derive the meteorological drought index (the SPI) in Chapter 7. Note that the KMA continuously provides weather records such as rainfall, temperature, wind speed and humidity across South Korea, while the K-water is more focused to observe hydrological data (e.g., runoff, rainfall and SM) for the purpose of dam operation.

3.2.2 Satellite SM data

3.2.2.1 ASCAT

The ASCAT onboard the METOP satellite crossing the Equator at 09:30 (descending orbit) and 21:30 (ascending orbit) local time was initially developed to monitor sea surface wind speed and direction using an active microwave remote sensing (Wagner et al., 2013). The ASCAT is a C-band radar operating at 5.3 GHz, and its SM retrieval algorithm was introduced by the Vienna University of Technology (TU Wien). Besides its initial purpose, the results of numerous validation studies carried out around the world have yielded clear evidence that the ASCAT also provides SM estimates with high reliability (Wagner et al., 2013). In addition, the ASCAT produces SM products with reasonable temporal resolution (revisit time: 1-3 days) and spatial resolution of 25-50 km (Figa-Saldaña et al., 2002). The ASCAT SM products can be obtained from either the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) or the H-SAF Products Download Centre. In this study, the ASCAT SM time series products (H109 Metop ASCAT DR2016) with a 12.5 km spatial resolution (resampled from a 25 km grid), which provides SM content in relative units ranging between 0% (driest condition) and 100% (wettest state), were collected from H-SAF (http://hsaf.meteoam.it; accessed on 28 July 2016). In Chapter 4, the ASCAT SM products obtained from both the ascending and descending passes are explored to examine the accuracy in terms of their orbits as well as temporal variation patterns.

3.2.2.2 AMSR2

AMSR2 onboard the GCOM-W1 satellite has been providing measurements of the surface SM since July 2012. Unlike the ASCAT, which uses active microwave remote sensing techniques, the AMSR2 is a passive microwave sensor, taking measurements at multiple frequencies to provide various hydrological parameters. The AMSR2 was developed to measure the brightness temperatures at seven different frequencies (including 6.9 GHz, 7.3 GHz, 10.65 GHz, 18.7 GHz, 23.8 GHz, 36.5 GHz and 89.0 GHz) and was initially designed to observe various parameters connected to the hydrological cycle, such as SM content, snow depth, precipitation and wind speed (Imaoka et al., 2010).

As a successor sensor to AMSR-E, which was in operation from 2002 to 2011, the basic theory of AMSR2 is almost the same as that of its predecessor (i.e., AMSR-E). Yet, the AMSR2 shows improvements compared to the AMSR-E; a 7.3-GHz channel was added to identify and address Radio Frequency Interference (RFI) signals, and AMSR2's antenna diameter was enlarged to 2 metres (AMSR-E's measures 1.6 metres) for better spatial resolution (JAXA, 2013; Wu et al., 2016). AMSR2 SM products can be obtained from two different algorithms (i.e., the JAXA (Koike, 2013) or Land Parameter Retrieval Method (LPRM; Owe et al., 2008)), and they can be downloaded from each distributor's website (https://gcom-w1.jaxa.jp for JAXA and http://gcmd.gsfc.nasa.gov for LPRM). Unlike the JAXA algorithm, which uses a 10.7 GHz channel, the LPRM approach generates SM products using the 6.9 (C-band), 7.3 (C-band) and 10.7 GHz (X-band). Before utilizing the AMSR2 SM product in this thesis, each dataset is compared to in-situ observations in Chapter 4. Based on the preliminary analysis, the JAXA algorithm showed the best agreement with in-situ observation in terms of the correlation coefficient. The results are discussed more in detail in Chapter 4. As a result, JAXA AMSR2 Level 3 (AMSR2) SM products (with 0.1° spatial resolution and volumetric terms (%)) are selected for further analysis in this thesis. In Chapter 5, the proposed downscaling scheme in this thesis is validated with respect to AMSR2 SM products.

3.2.2.3 ESA CCI

The ESA has released satellite-based SM products at a global scale by combining various individual active and passive microwave sensors (Dorigo et al., 2015). More specifically, the collected SM

products from individual active and passive sensors are blended by a spatio-temporal resampling and rescaling (i.e., the CDF matching) technique (Dorigo et al., 2015). The ESA CCI SM datasets are composed of active, passive and active plus passive products, covering the period 1978-2016. The ESA CCI SM products (v04.2) can be freely obtained from the distributor's web pages (https://www.esa-soilmoisture-cci.org/; accessed and downloaded on 17th September 2018). The ESA provides SM datasets to the public domain with different types of units (i.e., volumetric units (m^3/m^3) for the passive and combined product and per cent of saturation units (%) for active dataset). In terms of the active product, Active Microwave Instrument – Windscat (AMI-WS) and ASCAT are blended together. Similarly, the passive product is produced by SMMR, SSM/I, TRMM Microwave Imager (TMI), AMSR-E, WindSat, AMSR2 (successor to AMSR-E) and SMOS data. Figure 3-7 shows the timeline of input satellite sensors used for EAS CCI SM products. More details on the blending methodology as well as the evaluation study against ESA CCI SM can be found in the literature (Dorigo et al., 2017, 2015; Liu et al., 2012). In this thesis, the active-passive combined SM products with a spatial resolution of 0.25° are in particular chosen due to their higher temporal resolution and better accuracy compared to either the active or the passive retrievals (Liu et al., 2011). The obtained SM products are then incorporated into a hybrid RR modelling framework in Chapter 6.

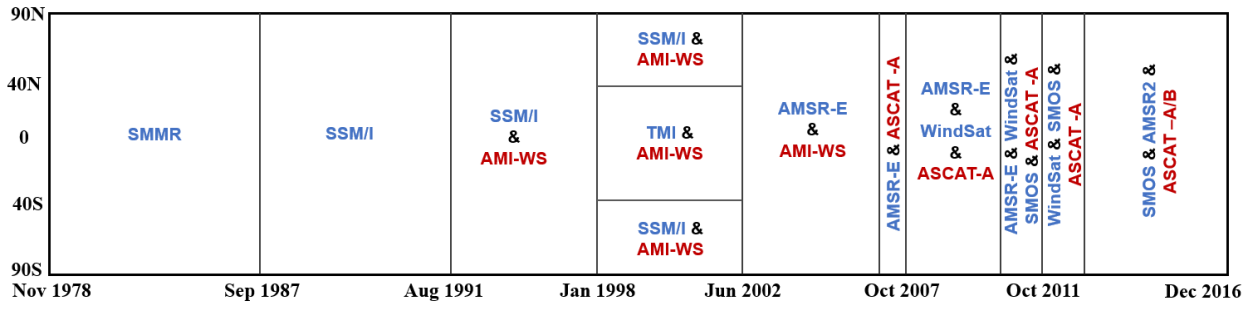


Figure 3-7 Timeline of input satellite sensors integrated into the ESA CCI products. This figure is presented based on Dorigo et al. (2017).

3.2.3 In-situ SM data

Similar to rainfall data, in-situ SM data collected in this thesis are also managed by two organizations (the KMA and K-water). The SM contents at depths of 10, 20, 40 and 50 cm have been measured by the KMA, while the K-water has provided SM observations at different

measurement depths (10, 20, 40, 60, 80 cm). A total of 12 sites across South Korea are selected in this thesis.

Figure 3-8 In-situ SM meter along with rain gauges in the Yongdam catchment.

SM data collected from the KMA are measured by using FDR sensors providing volumetric SM, while the K-water provides SM data in the Yongdam catchment by using TDR sensors. These observation datasets are assumed as the ground truth in assessing other sources of datasets such as

satellite, ERA-Interim and simulated SM. Figure 3-9 shows the location of the two SM networks used in this thesis. Here, KMA and YD represent (a) KMA SM network, and (b) K-water SM network, respectively.

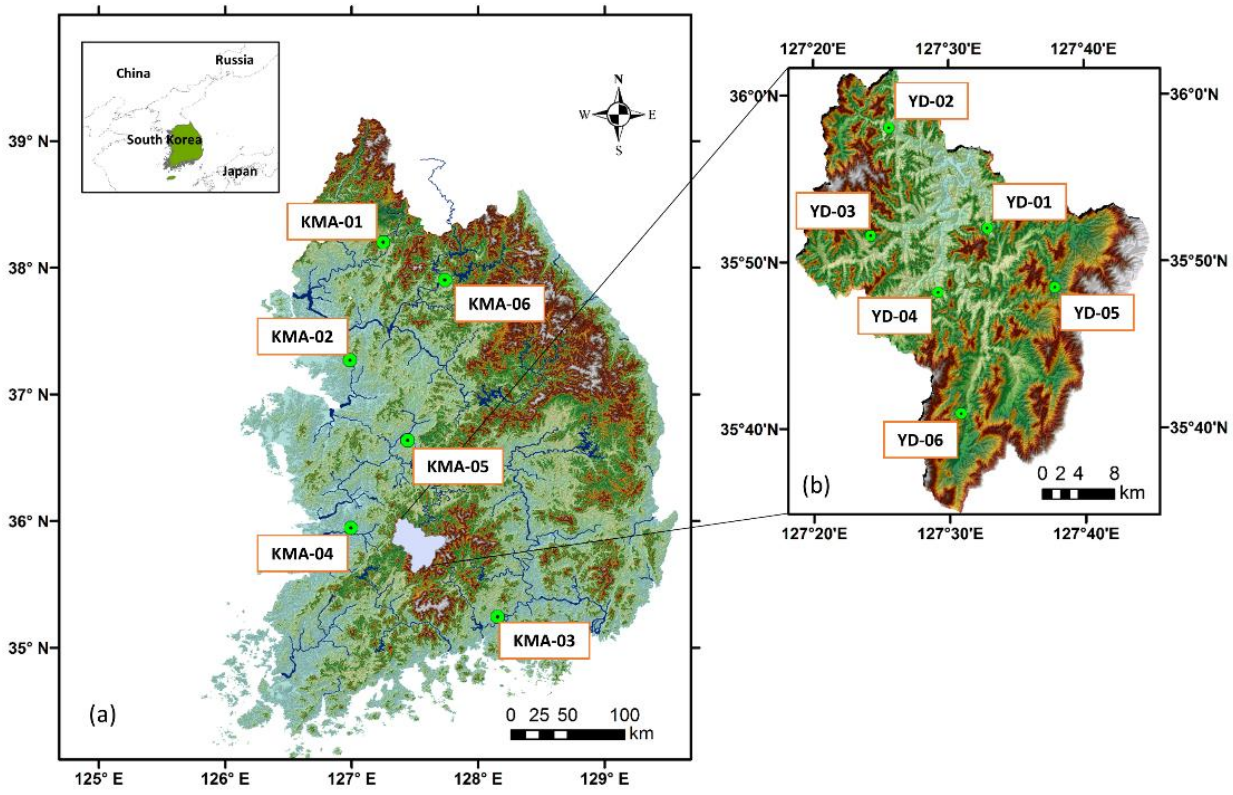


Figure 3-9 Locations of the two SM networks. Here, KMA and YD refer to KMA and K-water SM observation networks, respectively.

3.2.4 ERA-Interim SM data

Besides satellite and in-situ SM, a global SM data set from the European Centre for Medium-Range Weather Forecasts (ECMWF) are used in this thesis. ECMWF continuously releases global reanalysis SM datasets (hereinafter named ERA-Interim SM) with a high spatial resolution, daily at intervals of six hours in quasi-real time at four depths (i.e., 0–7, 7–28, 28–100 and 100–289 cm) (Albergel et al., 2012). The ECMWF provides ERA-Interim SM datasets with a spatial resolution of 25 km covering the period 1979 – present and they can be obtained from https://www.ecmwf.int/. As stated in Section 2.2, the accuracy of the ERA-Interim reanalysis data was globally explored by Albergel et al. (2012), revealing the robustness for various climate conditions with a reasonable

level of accuracy. The similar results were also achieved based on the preliminary analysis in the study area. This is, the Pearson correlation coefficients between the original ERA-Interim and the currently available in-situ SM data over all of South Korea are reasonably high, ranging from 0.60 to 0.75. In this thesis, 6-hourly SM data (0:00, 6:00, 12:00 and 18:00 UTC) whose centroid is nearest to the weather stations are first collected and then averaged to obtain a daily mean SM time series. The knowledge of SM at a given depth in the root zone is more meaningful hydrologically. Therefore the ERA-Interim SM data at the third layer (28–100 cm) are mainly used in this thesis for a drought analysis.

3.2.5 Summary of datasets

The SM data used in this thesis can be grouped into three subsets based on the sources of SM: ground-, satellite- and model-based SM. In terms of in-situ SM observations, the KMA and K-water have been providing ground-based measurements since 2013 and 2014, respectively. Although in-situ observations have become available over the past few years, collecting spatially and temporally appropriate data for hydrological applications is up to now significantly limited in South Korea. Nonetheless, SM data obtained from the observation networks are potentially useful, providing a reference for a validation procedure. In this respect, these in-situ observations are mainly used in this thesis for evaluating the performance of SM data obtained from other retrieval methods.

In Chapter 4, the AMSR2 and ASCAT satellite SM are assessed against in-situ observations and the scaling and filtering method are further applied to improve their applicability to practical applications. In a preliminary study, I attempted to evaluate SMOS and SMAP SM products. However, the number of available data acquired from both satellites was too small for their effective evaluation within the period investigated. It is found that observations at L-band are seriously perturbed by RFI (Colliander et al., 2017), and Asia and Europe together comprise the majority of RFI sources in the world (Oliva et al., 2012). In this respect, Zeng et al. (2015) have suggested that in Asia, which is the continent most contaminated by RFI, it is better to use other satellite sensors instead of the SMOS.

In Chapter 5, a stochastic downscaling scheme is introduced to spatially disaggregate the AMSR2 data, taking other hydrological variables (e.g., rainfall, temperature and in-situ observation) into

consideration as auxiliary predictors for SM simulation. The AMSR2 SM data are in particular selected owing to their relatively high temporal resolution compared to the ASCAT (Figure 3-10).

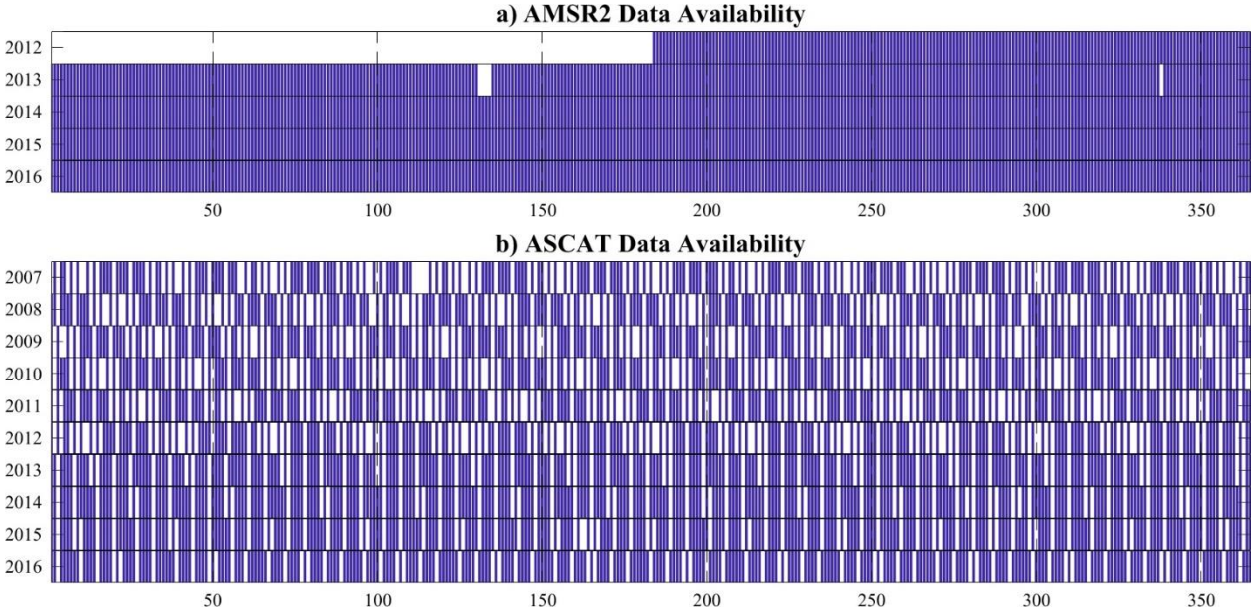


Figure 3-10 Data availability for a) AMSR2 and b) ASCAT, respectively. Here, the white bar represents missing data.

In terms of the ESA CCI and ERA-Interim datasets, ESA CCI SM data are incorporated into an RR modelling framework in Chapter 6, while ERA-Interim SM products are used for exploring the nature of drought presented in Chapter 7. The summary of the datasets used in this thesis is shown in Table 3-2.

Table 3-2 Summary of data information used in this thesis.

Chapter	Variable	Data Type	Data Source	Resolution	Scale
Ch. 4	In-situ SM	Point	KMA, K-water	12 stations	National
	ASCAT	Grid	H-SAF (H109 DR2016)	12.5 km	
	AMSR2	Grid	JAXA (Level3)	10.0 km	
Ch. 5	In-situ SM	Point	K-water	6 stations	Catchment
	AMSR2	Grid	JAXA (Level3)	10.0 km	
	Temperature	Point	KMA	1 station	
	Rainfall	Point	K-water	6 stations	
Ch. 6	In-situ SM	Point	KMA, K-water	12 stations	Catchment
	Rainfall	Point	K-water	6 stations	
	ESA CCI SM	Grid	ESA (v04.2)	25.0 km	
	Runoff	Point	K-water	1 station	
Ch. 7	In-situ SM	Point	K-water	6 stations	National
	ERA-Interim SM	Grid	ECMWF	25.0 km	
	Rainfall	Point	KMA	55 stations	

CHAPTER 4 [1]Assessment of remotely sensed soil moisture

4.1 Motivation

SM retrieved from satellite observations has become available at a global scale with relatively high spatial-temporal coverage, and satellite-derived SM can be useful data sources where in-situ measurements are scarce or not available, such as in South Korea. Among many satellite retrievals, this chapter focuses on two sensors (AMSR2 and ASCAT) based on practical issues (i.e., data availability), as well as the results of previous studies (Zeng et al., 2015) as mentioned in Section 3.2. Prior to the actual use of satellite SM products, it is necessary to evaluate the accuracy and reliability by comparing them with reference datasets. Satellite sensors could produce different levels of efficiency for different parts of the world as the performance of satellite retrieval is easily contaminated by surface roughness and vegetation, which potentially varies from region to region (Wagner et al., 2013). In this sense, this chapter begins with exploring the accuracy of the original satellite SM retrievals in terms of their orbits and temporal variation patterns.

Satellite-retrieved SM is representative of a topsoil layer, while the RZSM plays a more important role in hydrological applications such as drought detection and agricultural modelling because it presents plant-available water (González-Zamora et al., 2016). The exponential filter method is adopted to derive the RZSM from original satellite SM retrievals, making it more applicable to practical issues. However, there still exist systematic biases between two different data sources (i.e., in-situ and satellite SM), which also constrains the direct use of satellite SM. In this sense, bias correction methods are particularly employed as a pre-processing tool for original satellite SM products. Most of the conventional CDF matching schemes are conducted based on predefined temporal scales. In this chapter, the proposed CDF matching approach is applied, not only to address the inherent systematic mismatch between the two datasets but also to select an optimal temporal resolution.

---

[1]Moonhyuk Kwon and Dawei Han. "Assessment of Remotely Sensed Soil Moisture Products and their Quality Improvement." *Journal of Hydro-environment.* Doi: 10.1016/j.jhydrol.2017.12.015.

In other words, as well as the conventional CDF matching method that uses the whole record of the investigation period, this chapter explores the performance of CDF matching methods on a different temporal resolution basis to select an ideal combination: monthly (12 groups), seasonal (4 groups), and growing and non-growing (2 groups). The performance of each bias-correction group is then validated through a cross-validation procedure.

Given this background, this chapter aims to answer the following questions:

(1) What is the reliability of SM retrievals from satellite sensors (ASCAT and AMSR2) and how does their performance in South Korea differ from that of other parts of the world? Does acquisition time (i.e., ascending and descending overpass) affect the quality of satellite SM retrievals?

(2) How could the applicability of satellite SM be improved? Is it desirable to apply the SWI approach to derive RZSM from the surface, and are there any limitations to the utilisation of the SWI method?

(3) Is the CDF matching method a useful scheme in mitigating the systematic biases between in-situ and satellite data? Do the different combinations of temporal periods affect results?

4.2 Methodology

4.2.1 Study area and datasets

As described in Chapter 3, the observed SM data are obtained from two organization (the KMA and K-water). Figure 3-9 shows the study areas along with twelve in-situ SM observation stations throughout South Korea and the main characteristics of SM sites are presented in Table 4-1. The satellite SM product sets retrieved from both ASCAT (active microwave sensor), and AMSR2 (passive microwave sensor) are compared with the in-situ SM observations (as ground truth) to evaluate their performance. The satellite pixel values whose centroids are located nearest to each ground observation site are extracted for this study. Owing to differences in spatial-temporal resolutions as well as observation depths between satellite and point measurements, satellite data are usually scaled and/or filtered before their utilization for actual applications (Scipal et al., 2008).

In the first step, given that SM estimates are provided by different units (volumetric terms for both in-situ and AMSR2, and relative SM for ASCAT), I normalised all the data by using the maximum and the minimum values over the investigation period. Both satellite data and in-situ observations have the same maximum and minimum values by employing the normalising method.

Table 4-1 Main characteristics of the study sites. Here, water fraction indicates the area ratio of wetlands plus open water surfaces within ASCAT pixel (12.5 km).

Site	Elevation (m a.s.l)	Longitude (º)	Latitude (º)	Annual rainfall (mm/year)	Observation depth (cm)	Land use	Water ratio (%)
KMA-01	181.0	127.25	38.20	1,179	10, 20, 30, 50	Forest	1.6
KMA-02	33.6	126.99	37.27	1,007	10, 20, 30, 50	Agriculture	3.2
KMA-03	22.0	128.15	35.24	1,397	10, 20, 30, 50	Forest	4.5
KMA-04	15.0	126.99	35.95	1,095	10, 20, 30, 50	Agriculture	3.9
KMA-05	56.4	127.44	36.63	970	10, 20, 30, 50	Agriculture	2.2
KMA-06	76.8	127.74	37.90	1,058	10, 20, 30, 50	Forest	9.1
YD-01	313.0	127.55	35.87	1,011	10, 20, 40, 60, 80	Forest	2.0
YD-02	330.0	127.43	35.97	1,111	10, 20, 40, 60	Forest	0.7
YD-03	396.0	127.40	35.86	1,108	10, 20, 40, 60	Forest	0.4
YD-04	334.0	127.49	35.80	1,043	10, 20, 40, 60, 80	Forest	1.4
YD-05	453.0	127.63	35.81	956	10, 20, 40, 60, 80	Forest	0.6
YD-06	409.0	127.51	35.68	1,071	10, 20, 40, 60, 80	Forest	0.7

4.2.2 Filtering technique

Satellite-retrieved SM is representative of a topsoil layer (i.e., satellite-based SM estimates have inherent limitations in capturing the variation of the RZSM), while the RZSM is more readily applicable to be incorporated into hydro-meteorological models (Brocca et al., 2012; Dharssi et al., 2011). In this sense, one popular semi-empirical approach, the exponential filter technique is adopted to derive the RZSM from near-surface observations. As stated in Section 2.3, the performance of the SWI is substantially dependent on the characteristic time (T). Here, the T value is determined by optimising the correlation coefficient (r) between the SWI and in-situ observation. In other words, the T value corresponding to the highest correlation between the SWI and in-situ observation is considered as the optimal characteristic time length (T) for each SM observation stations. The derived SWI is then compared with in-situ SM with respect to different observation depths along with depth averaged SM contents (θ_{0-60}). In this study, the profile SM referring to

depth-weighted mean SM content between the surface and a 60cm soil depth is computed as follows:

$$\theta_{0-60} = \frac{\theta_i \cdot d_i + \frac{\theta_i+\theta_{i+1}}{2}(d_{i+1}-d_i) + \frac{\theta_{i+1}+\theta_{i+2}}{2}(d_{i+2}-d_{i+1}) + \frac{\theta_{i+2}+\theta_{i+3}}{2}(d_{i+3}-d_{i+2})}{d_{i+3}} \quad (4\text{-}1)$$

where d_i(cm) represents the *i*-th depth of measurement from the top layer, and θ_i(%) is the SM obtained from the *i*-th depth. In the case where measurements at the 60 cm depth are not available, the values at the 60 cm depth were replaced by SM measurements at the 50 cm depth.

4.2.3 Scaling technique

As stated in Section 2.3, the mismatch in spatial scale and measuring depth between satellite-based retrievals and ground-based observations are likely to cause inevitable systematic biases. In this chapter, the CDF matching approach is in particular considered to tackle systematic biases between different data sources. Through this method, the biased data (i.e., original satellite SM data) are mapped to the same probability value as that of observations (Figure 2-3).

In this chapter, the CDF matching is implemented using a polynomial fit. Namely, the CDF of the two datasets (i.e., the satellite-derived SWI and in-situ observations) is firstly displayed, and then the differences corresponding to the CDF of each ranked data are computed. Next, the observation operator is derived based on a polynomial fit, which allows defining site-specific parameters. To be specific, the parameters of the polynomial equation are estimated from one subset, and the derived parameters are then exploited to the remaining data set for validation. In addition, I test the performance of observation operators based on four different temporal groups. More groups are likely to result in reducing error, while using too many groups can lead to the overfitting issue. To avoid overfitting, the parameters obtained from the calibration period are tested for validation.

4.2.4 Performance Indices

The performance and accuracy of satellite SM retrievals are assessed by comparing them against ground-based observations that are regarded as reference SM values. For this study, four statistical indicators (i.e., correlation coefficient (*r*), root mean square error (RMSE), unbiased RMSE (ubRMSE) and bias) are computed to quantify the level of accuracy, which are commonly adopted in the field of remote sensing (Zeng et al., 2015). Here, for *N* discrete datasets of two variables

(i.e., satellite SM retrieval (θ_s) and in-situ observation (θ_n)), the Pearson correlation coefficient (r) is used to examine temporal pattern similarity between two datasets, given by:

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N}(\theta_{s,i} - \overline{\theta_s})(\theta_{n,i} - \overline{\theta_n})}{\sigma_S \, \sigma_n} \quad (4\text{-}2)$$

where σ_S and σ_n represent the standard deviation of satellite and in-situ SM, respectively. The overbar indicates the averages over the entire investigation period. In addition to the correlation coefficient, root mean square error (RMSE) and unbiased root mean square error (ubRMSE) are used for the validation of satellite SM products. RMSE and ubRMSE ($ubRMSE^2 = RMSE^2 - bias^2$) are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{n,i} - \theta_{s,i}\right)^2} \quad (4\text{-}3)$$

$$ubRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(\theta_{s,i} - \overline{\theta_s}\right) - \left(\theta_{n,i} - \overline{\theta_n}\right)\right]^2} \quad (4\text{-}4)$$

4.3 Results and discussion

In this section, I evaluate the accuracy and reliability of satellite-based SM products. The satellite SM products retrieved from both ASCAT and AMSR2 are compared against ground-based measurements collected from 12 different sites, over a three-year period for KMA sites (2013-2015), and a two-year period for YD sites (2014-2015). As mentioned in Section 3.3, in-situ SM data used in this thesis are managed by two organizations (the KMA and K-water).

4.3.1 Assessment of original satellite SM data

Prior to evaluating the satellite-based SM products, I first attempt to explore the performance of SM retrieval algorithms (for AMSR2). Here, I assess each retrieval algorithm by comparing it with in-situ data measured at a depth of 10 cm. In terms of the LPRM algorithms, there is no significant improvement in accuracy by applying different frequencies (X, C1 and C2 band), with mean *r* values ranging between 0.14 and 0.17 for 12 observation sites (Table 4-2). In terms of the mean *r* value, AMSR2-JAXA shows a similar accuracy level, but satellite SM data with a higher spatial-

temporal resolution can be obtained by using JAXA algorithm (10 km for JAXA and 25 km for LPRM). Taking this advantage into account, the AMSR2 SM data derived from JAXA algorithm are hereinafter used for further studies.

Table 4-2 Comparison of different retrieval algorithms for AMSR2 SM products.

Algorithm	Frequency	mean r	mean RMSE	mean Bias	max r	Min r
JAXA	10.7	0.17	0.45	0.38	0.29	0.02
LPRM (X)	10.7	0.14	0.29	-0.05	0.46	-0.20
(C1)	6.9	0.15	0.38	-0.24	0.28	0.05
(C2)	7.3	0.17	0.36	-0.12	0.28	0.01

As for polar orbit satellites, SM products are provided at different acquisition times (i.e., ascending and descending overpasses). It is necessary to examine the performance associated with their overpass time. The night-time retrievals are generally expected to have higher accuracy than the daytime products since the geophysical conditions are more favourable during night-time (Kim et al., 2015; Zeng et al., 2015). On the other hand, there is also a positive effect over daytime in that the canopy is more transparent and drier during daytime (Brocca et al., 2011). Here, daytime refers to the ascending overpass for AMSR2 (1:30 pm) and descending overpass of ASCAT (9:30 am), and vice versa for the night-time retrieval. For this study, in-situ observations measured at 10 cm depth corresponding to the satellite overpass times are used to evaluate the performance with respect to orbit direction. As it can be seen from Figure 4-1, the descending retrieval for ASCAT is shown to be superior to the ascending one, while no significant discrepancy can be found for AMSR2.

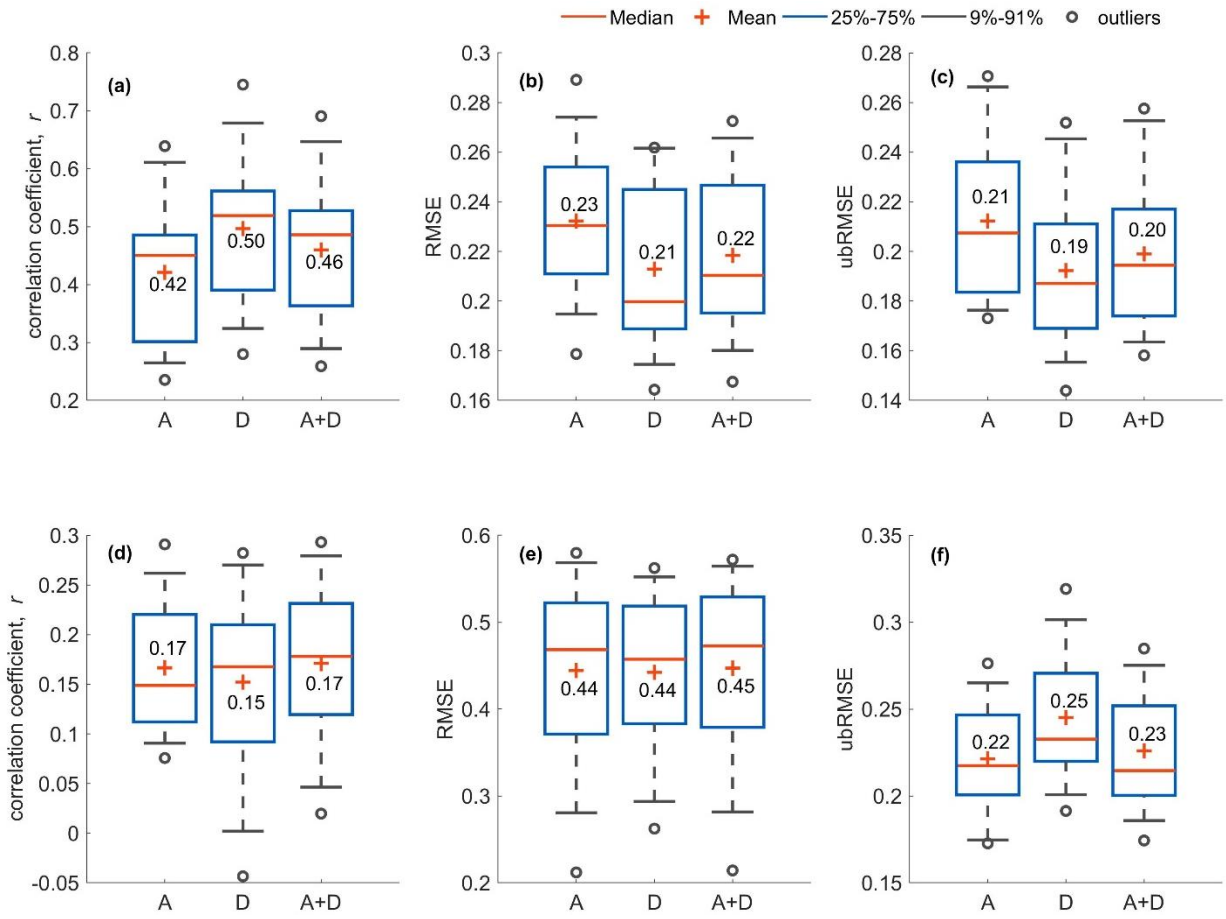


Figure 4-1 Boxplots of the correlation coefficient (r), RMSE and ubRMSE: (a-c) for ASCAT and (d-f) for AMSR2. Here, the x-axis indicates satellite orbits; (A) and (D) correspond to the ascending and descending overpasses, respectively. (A+D) refers to the aggregation of the ascending and descending overpasses.

The results for ASCAT are in accordance with the findings by Griesfeller et al. (2016) who obtained mean *r* values for Norway equal to 0.72 for the descending orbit (daytime) and 0.68 for the ascending orbit (night-time), namely the descending overpass outperforms the ascending overpass. Interestingly, they also found descending retrievals (night-time) to be in better agreement with in-situ observations for AMSR-E. In contrast, Zeng et al. (2015) obtained a higher *r* value for the ascending orbit in China (0.788 for night-time and 0.885 for daytime). The abovementioned studies indicate that the accuracy of SM data with respect to satellite orbit is highly location-dependent: SM products from the satellite can be affected not only by the orbits but also by other various factors such as soil texture, topography, land cover, and climate. For instance, the *r* values

for the KMA01 site are equal to 0.64 for the ascending overpass, 0.75 for the descending overpass, and 0.69 for the ascending plus descending overpasses (Figure 4-2). Compared to the descending overpass, the combination of ascending and descending overpasses shows a negligible decrease in performance in terms of *r* value. Furthermore, the combination of ascending and descending overpasses increases the temporal data coverage to 91% (N: 991) of date for the investigation period without consideration of any interpolation schemes (Figure 4-2c). In this study, both the ascending and descending products are used to obtain higher temporal coverage, which may aid in providing more robust results by increasing the amount of data analysed. For this reason, both passes were commonly used in previous studies (Brocca et al., 2011; Kolassa et al., 2016).

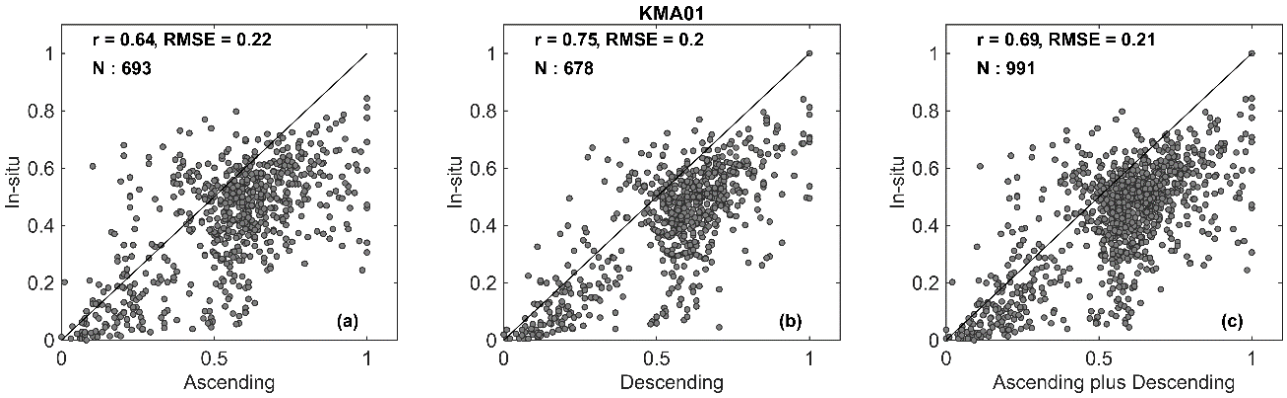


Figure 4-2 Statistical scores (*r* and RMSE) between ASCAT SM and site-specific data sets for the KMA01 site. Here, N indicates the number of data pairs.

To examine how SM products perform seasonally and annually, a time series comparison of the different data sources from two sites is graphically presented in Figure 4-3. The seasonal variation is substantial over the study sites, displaying the characteristic of monsoons. The ASCAT SM retrievals appear to overestimate in-situ data, while AMSR2 generally underestimates the SM. The results obtained are consistent with previous studies (Cho et al., 2015; Kim et al., 2015; Zeng et al., 2015). They also found that the AMSR2 retrievals tend to underestimate in-situ SM with unrealistically high values responding to precipitation events and the lack of temporal dynamics.

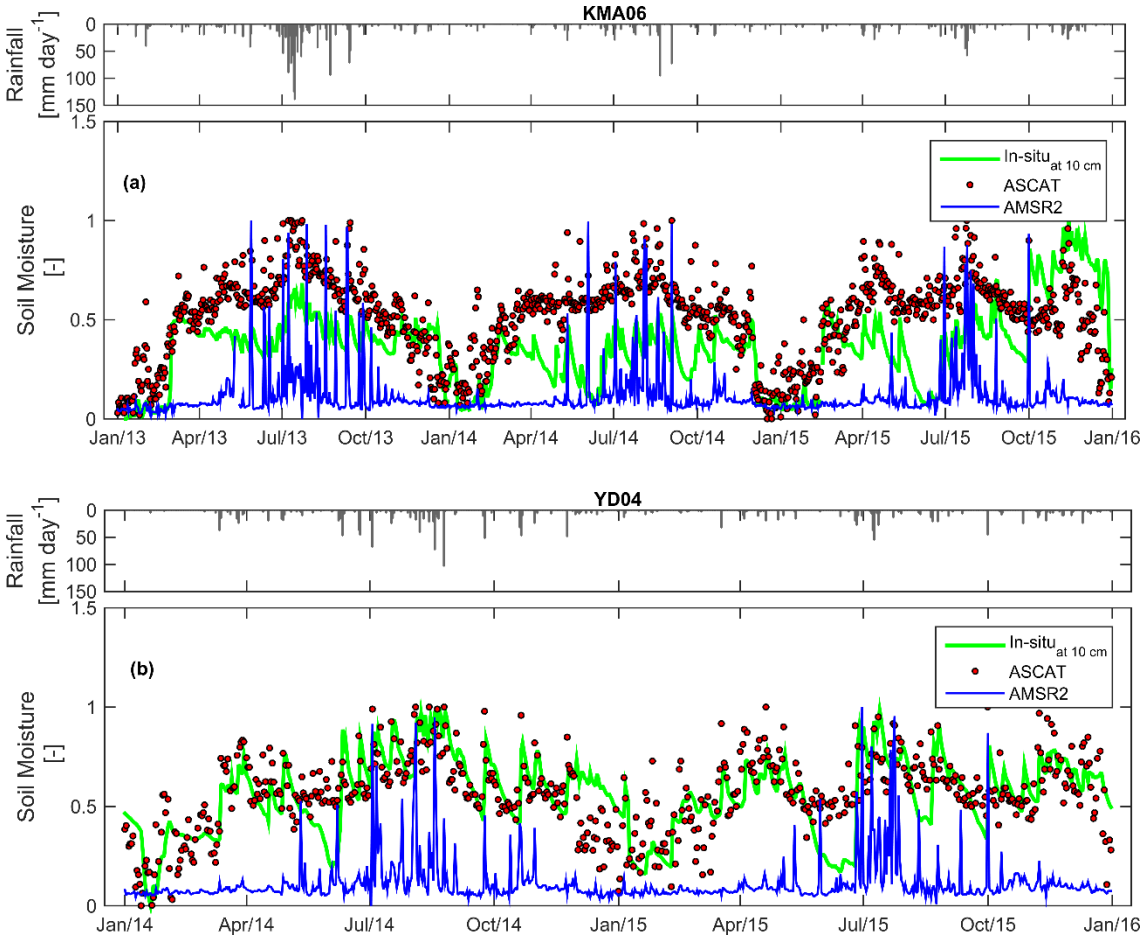


Figure 4-3 Samples of time series comparison of SM products (ASCAT and AMSR2) with in-situ observations. The bar graph indicates rainfall.

4.3.2 Assessment of ASCAT data

4.3.2.1 Results after filtering

The microwave-based ASCAT products are representative of a very shallow soil layer (Brocca et al., 2011), whereas they are compared with in-situ observations measured greater than a depth of 10 cm. Moreover, the RZSM is a more important variable for many hydrological applications. In this regard, a recursive exponential filter method that allows estimating the RZSM from the surface measurement is employed. Then, the derived SWI from ASCAT surface SM products are compared with the in-situ SM observations at different depths along with the SM profile from surface to 50 cm depth ($d_{0-50\,cm}$). Here, correlation coefficient (r) is used for the selection of the optimal T, based on the fact that it is more meaningful to capture the temporal behaviour of SM

rather than the absolute value for many hydrological applications (González-Zamora et al., 2016). Table 4-3 shows the statistical performance between the ASCAT SWI and in-situ observations measured at different depths at 12 sites. The mean *r* values are 0.54, 0.52, 0.51 and 0.47 at 10, 20, 30 and 50 cm depth, respectively, and a slightly higher *r* value (0.58) is obtained from the SM profile (0-50 cm).

In all the observation depths, the results show improved temporal correlations with in-situ observations, indicating that the SWI method can reproduce the behaviour of the RZSM. However, the relatively large differences in *r* values among the sites are found owing to systematic biases between the original satellite and in-situ observations. In terms of the mean RMSE, the figures are equal to 0.19, 0.21, 0.22, and 0.25 at the depths of 10, 20, 30 and 50 cm respectively, confirming a better performance of the SWI at the shallow soil layer. The differences in mean ubRMSE for each observation depth, however, are negligible ranging from 0.16 to 0.18. Considering relatively large differences between the ubRMSE and the RMSE (i.e., there remain systematic biases between in-situ and satellite SM dataset), it can be argued that bias reduction techniques should be employed to improve the accuracy of satellite retrievals with respect to in-situ observations.

The characteristic time length (T), representing the SM travel time from the surface, increases as the depth increases (3.1 days for 10 cm and 8.3 days for 50 cm), which is in line with the assumption of the SWI approach. The optimal T value for 0–50 cm shows similar results to those obtained for 10 cm, which indicates that the SM stored in the topsoil layer have more influence on the SM profile (0-50 cm). For SM profile (0-50 cm), one of the leading factors impacting the satellite SM is the ratio of open water surface within the pixel: the KMA01 site with the smallest rate of open water surface (1.5%; Table 4-1) has the best *r* value of 0.83, but the KMA06 site with the highest proportion (9.1%) shows the lowest *r* value of 0.53. One possible explanation is that the concept of a microwave remote sensing technique is based on the difference in dielectric constant between water and dry soil. Thus, a higher fraction of water body within a pixel size deteriorates the accuracy in monitoring surface SM.

However, in the case of YD sites, the ratio of open water surface (< 2.0%) is much smaller than that of KMA sites, and there is no significant difference in *r* value according to the ratio of open water surface. However, some of the observation sites show surprising results of T values being smaller for the deeper soil layer. For instance, the optimal T value at the YD03 site appears to be

inconsistent with the model assumption (i.e., 3.7 days for 10 cm depth and 1.5 days for 60 cm depth, respectively). A feasible explanation is presented in Figure 4-4, showing an example of the dynamic range of the SWI with respect to T values. Here, it is clear that as the T value increases, the time series of the ASCAT SWI becomes smoother (Figure 4-4a). In other words, the lower dynamic range with a larger T value is generally expected to be representative of SM contents at a deeper soil layer rather than a topsoil layer.

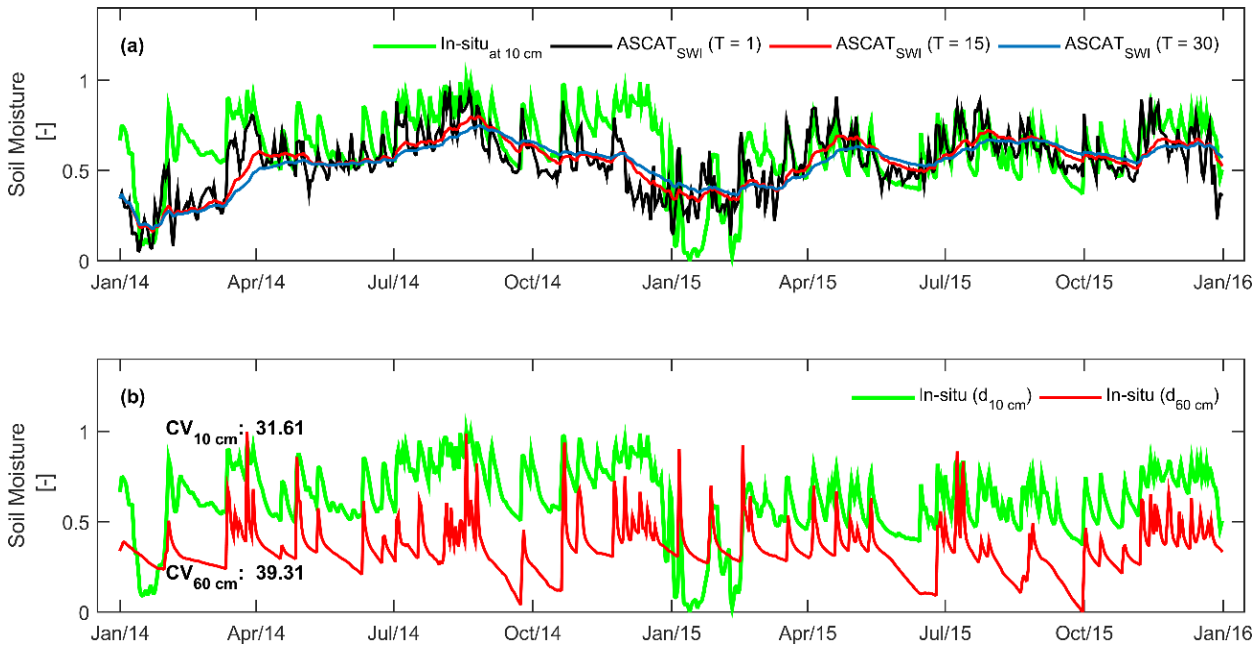


Figure 4-4 (a) Time series plot showing in-situ SM observations and ASCAT SWI from the YD03 site with respect to different T values (1, 15 and 30). (b) Time series plot showing in-situ observations at different observation depths along with the coefficient of variation (CV).

Interestingly, in this specific case, in-situ SM time series at a depth of 60 cm shows rather larger temporal variability compared with that measured at 10 cm depth, with a coefficient of variation (CV) equal to 31.61 for 10 cm and 39.31 for 60 cm (Figure 4-4b). The results are against the basic concept of the exponential filter method that assumes the SM content integrated over the deeper layers, thus exhibiting less variation than in the topsoil layer (González-Zamora et al., 2016). However, at some of the in-situ observations in this study, SM contents at the lower layer tend to respond more rapidly to rainfall, which may be caused by many uncertain factors. This abnormal SM variation at the deeper soil layer might be attributed to a preferential flow, creating an uneven and often rapid movement of water in the soil (Paquette et al., 2016). Nonetheless, it should be

noted that although the SWI approach is unlikely to capture short-term fluctuations that may occur in the root-zone in a particular area, the SWI method is a useful tool to reproduce the temporal dynamics of the RZSM.

4.3.2.2 Results after scaling

The CDF matching method is widely used in many hydrological applications to remove the systematic biases between two data sets. Here, the CDFs of the derived SWI are matched with those of in-situ observations at each site. The CDF matching method, in this study, is used to derive an observation operator through the third-order polynomial fit that has also been used in previous studies (e.g., Drusch et al., 2005; Han et al., 2012). The aim of using an observation operator is to define a set of parameters that are suitable for further use. In this study, besides the conventional CDF matching method that uses the whole record of the investigation period (QM1), I explore the performance of the CDF matching method on a different temporal resolution basis: monthly (12 groups; QM2), seasonal (4 groups; QM3) and growing and non-growing (2 groups; QM4). To be specific, the CDF matching method is built and validated for four different temporal groups: 1) the entire period of investigation, 2) monthly, 3) seasonal (spring (Mar-May), summer (Jun-Aug), fall (Sep-Nov) and winter (Dec-Feb)), and 4) growing (Apr-Sep) and non-growing seasons.

The proposed CDF matching approach is first tested to select an optimal temporal resolution in terms of statistical scores. For the sake of brevity, the results obtained at 10 cm are solely presented. Taylor diagram is displayed in Figure 4-5, illustrating the statistical metrics of the comparison between in-situ observations and satellite retrievals with respect to the aforementioned temporal groups. Compared to the result obtained from ASCAT SWI (Table 4-3), it is clear that the ASCAT SWI-CDFs present enhanced performance scores, with the exception of QM1. To be specific, QM1 shows a relatively low range of correlations with most values being less than 0.77 (mean r = 0.54). On the other hand, the mean r values increase from 0.54 (ASCAT SWI) to 0.78, 0.77 and 0.78 for QM2, QM3 and QM4 respectively. As for ubRMSE values, they also generally show improved results, though not as significant as r values.

Table 4-3 Comparison of ASCAT SWI with different observation depths (r: correlation coefficient, RMSE: root mean square error, T: characteristic time length (days)).

Site	D_{10cm}				$D_{20\ cm}$				$D_{30\ cm}$				$D_{50\ cm}$				$D_{0\text{-}50\ cm}$			
	r	RMSE	ubRMSE	T	r	RMSE	ubRMSE	T	r	RMSE	ubRMSE	T	r	RMSE	ubRMSE	T	r	RMSE	ubRMSE	T
KMA01	0.74	0.19	0.15	2.1	0.69	0.33	0.14	2.5	0.73	0.38	0.13	4.7	0.71	0.23	0.15	16.7	0.83	0.24	0.11	2.5
KMA02	0.42	0.22	0.15	1.7	0.70	0.13	0.12	4.1	0.67	0.15	0.14	4.3	0.66	0.20	0.16	3.9	0.71	0.14	0.13	2.9
KMA03	0.44	0.19	0.19	1.7	0.47	0.17	0.17	2.7	0.46	0.21	0.18	4.7	0.63	0.22	0.14	4.5	0.62	0.15	0.14	2.1
KMA04	0.59	0.16	0.15	2.3	0.63	0.24	0.18	5.9	0.60	0.22	0.18	6.9	0.39	0.29	0.19	6.3	0.63	0.17	0.17	3.1
KMA05	0.63	0.18	0.17	2.9	0.65	0.14	0.15	5.7	0.68	0.16	0.12	19.9	0.66	0.23	0.11	19.9	0.70	0.13	0.14	4.1
KMA06	0.53	0.25	0.18	4.3	0.61	0.18	0.18	11.3	0.51	0.21	0.14	19.9	0.21	0.43	0.14	19.9	0.53	0.26	0.15	8.1
YD01	0.65	0.16	0.13	4.3	0.53	0.23	0.17	3.3	0.45	0.29	0.20	2.9	0.74	0.22	0.17	5.9	0.59	0.20	0.16	2.9
YD02	0.31	0.25	0.24	2.3	0.22	0.26	0.24	2.1	0.25	0.23	0.22	1.7	0.06	0.23	0.22	1.3	0.28	0.22	0.22	1.3
YD03	0.60	0.18	0.16	3.7	0.47	0.19	0.17	2.5	0.39	0.20	0.20	3.1	0.29	0.27	0.17	1.5	0.55	0.17	0.16	2.1
YD04	0.76	0.12	0.14	5.5	0.66	0.15	0.15	6.7	0.76	0.18	0.14	8.9	0.68	0.21	0.15	9.3	0.76	0.16	0.15	6.5
YD05	0.46	0.20	0.20	3.3	0.30	0.23	0.22	3.7	0.17	0.24	0.24	4.9	0.12	0.27	0.27	5.5	0.34	0.22	0.20	3.3
YD06	0.39	0.22	0.22	3.1	0.32	0.23	0.23	3.3	0.42	0.21	0.18	4.7	0.44	0.24	0.20	5.3	0.41	0.21	0.20	3.1
Average	0.54	0.19	0.17	3.1	0.52	0.21	0.18	4.5	0.51	0.22	0.17	7.2	0.47	0.25	0.17	8.3	0.58	0.19	0.16	3.5

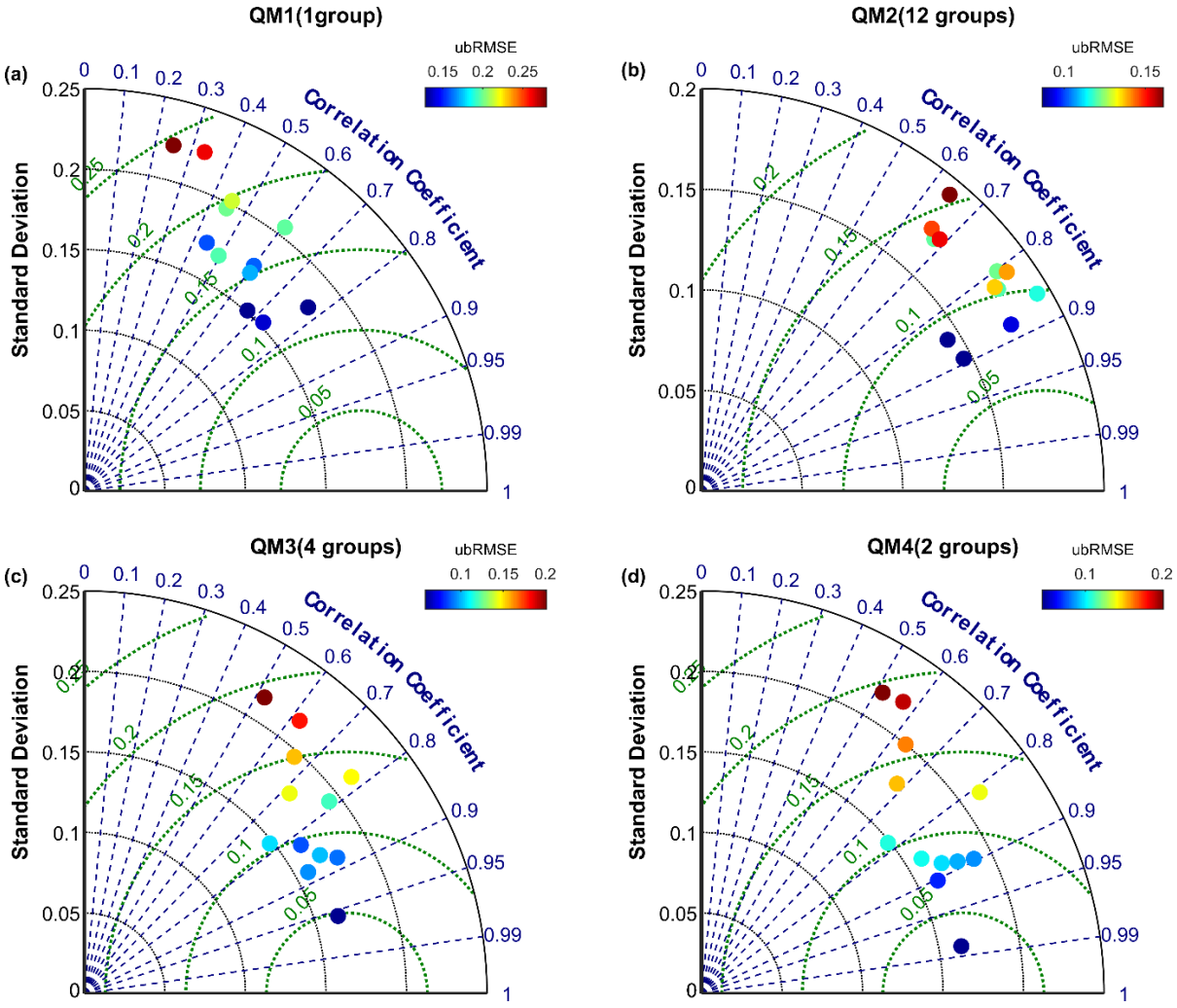


Figure 4-5 Taylor diagram representing the statistics between the in-situ observations measured at 10 cm depth and ASCAT SWI-CDF at 12 sites.

To further ensure the applicability of the observation operators, I partitioned the datasets into two subsets. The datasets of ASCAT SWI are initially grouped based on temporal resolution. Then, the established parameters of the polynomial equation for the calibration period are validated for the remaining datasets. The results of the CDF matching during calibration and validation period are presented in Figure 4-6. The observation operators behave differently between calibration and validation phases depending on temporal resolutions. The observation operators in calibration outperform those in validation phases. In terms of the correlation coefficient, the observation operator derived from the QM1 group shows a clearly worse performance compared to other temporal groups. Although both QM2 and QM3 display almost equally robust performances in statistical scores for calibration periods, the results obtained from the validation period show that

the highest mean *r* values are observed when the datasets are grouped on the basis of growing and non-growing seasons (QM4). The similar results are generally found with respect to the RMSE and ubRMSE.

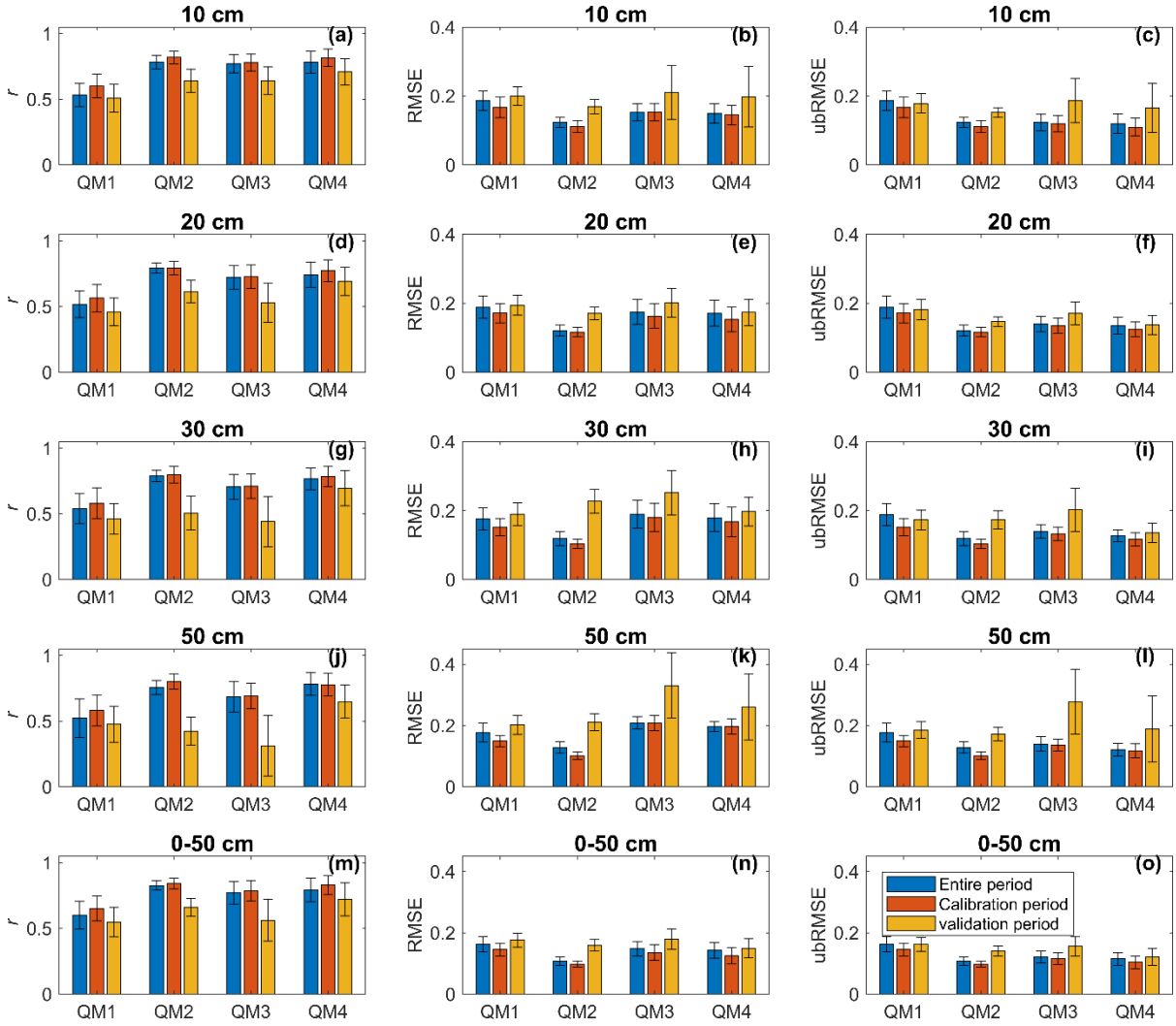


Figure 4-6 Statistics of the correlation coefficient (*r*), RMSE, and ubRMSE. Here, the error bar indicates 95% confidence interval.

4.3.3 Assessment of AMSR2 data

The AMSR2 SM products are evaluated against ground SM observations with the same procedure as the ASCAT: the scaling and filtering methods are also applied to assess and improve their performance.

Table 4-4 Comparison of AMSR2 SWI with different observation depths (r: correlation coefficient, RMSE: root mean square error, T: characteristic time length (days)).

Site	D_{10cm}				$D_{20\ cm}$				$D_{30\ cm}$				$D_{50\ cm}$				$D_{0\text{-}50\ cm}$			
	r	**RMSE**	**ubRMSE**	**T**	***r***	**RMSE**	**ubRMSE**	**T**	***r***	**RMSE**	**ubRMSE**	**T**	***r***	**RMSE**	**ubRMSE**	**T**	***r***	**RMSE**	**ubRMSE**	**T**
KMA01	0.49	0.32	0.17	7.3	0.51	0.16	0.11	7.1	0.64	0.10	0.08	19.1	0.72	0.28	0.16	17.7	0.70	0.18	0.10	13.9
KMA02	0.39	0.17	0.14	29.9	0.46	0.26	0.15	5.3	0.46	0.35	0.16	5.1	0.46	0.23	0.19	4.3	0.49	0.26	0.16	5.1
KMA03	0.29	0.37	0.21	2.7	0.11	0.37	0.23	2.5	0.00	0.31	0.25	2.3	0.19	0.24	0.21	3.5	0.18	0.32	0.21	2.7
KMA04	0.16	0.41	0.20	2.7	0.14	0.51	0.24	5.5	0.14	0.48	0.24	6.1	0.18	0.56	0.21	7.9	0.16	0.42	0.24	3.9
KMA05	0.10	0.44	0.24	2.7	0.17	0.47	0.19	29.9	0.25	0.52	0.16	29.9	0.44	0.24	0.14	29.9	0.21	0.43	0.18	29.9
KMA06	0.40	0.31	0.18	29.9	0.46	0.45	0.21	29.9	0.61	0.43	0.15	28.3	0.69	0.50	0.16	29.9	0.59	0.41	0.17	28.9
YD01	0.54	0.55	0.14	7.7	0.50	0.33	0.16	6.3	0.38	0.32	0.20	5.5	0.61	0.38	0.22	21.7	0.49	0.38	0.17	6.7
YD02	0.23	0.56	0.23	3.1	0.19	0.55	0.23	3.1	0.18	0.45	0.20	2.7	0.07	0.49	0.17	3.5	0.20	0.53	0.21	3.1
YD03	0.42	0.53	0.18	11.5	0.34	0.53	0.17	4.3	0.36	0.49	0.19	5.9	0.29	0.27	0.14	3.1	0.40	0.42	0.17	5.1
YD04	0.62	0.47	0.17	18.9	0.58	0.48	0.17	23.1	0.71	0.38	0.18	29.9	0.63	0.34	0.17	25.7	0.66	0.41	0.18	25.7
YD05	0.41	0.49	0.19	6.5	0.26	0.48	0.21	7.3	0.20	0.48	0.20	12.7	0.15	0.47	0.24	12.3	0.31	0.45	0.19	9.3
YD06	0.25	0.50	0.23	4.1	0.17	0.51	0.23	4.3	0.22	0.37	0.17	8.1	0.20	0.36	0.20	10.5	0.22	0.45	0.20	5.3
Average	0.36	0.43	0.19	10.6	0.33	0.42	0.19	10.7	0.34	0.39	0.18	13.0	0.39	0.36	0.18	14.2	0.38	0.39	0.18	11.6

4.3.3.1 Results after filtering

In the same manner as the ASCAT SM products, I derive the AMSR2 SWI using the exponential filter, and then the derived RZSM at each observation depth is compared with in-situ observations. Table 4-4 shows the statistical scores describing the degree of agreements between the AMSR2 SWI and in-situ observations measured at different depths. The average *r* values are equal to 0.36, 0.33, 0.34, 0.39, and 0.38 at 10, 20, 30, 50, and 0-50 cm depth, respectively, and a slightly higher mean *r* value is obtained from SM profile (0-50 cm). The mean RMSE for each observation depth ranges from 0.36 to 0.43, and the mean ubRMSE is from 0.18 to 0.19. The performance scores for AMSR2 are slightly lower than those obtained by ASCAT SWI. This is attributed to the discrepancy in the correlation of original AMSR2 data. It is interesting to note that the characteristic time (T) of the exponential filter is longer than that of ASCAT, with the average value of 10.6 days for AMSR2, and 3.1 days for ASCAT at 10 cm. The results support the findings of the previous studies, namely the optimal T highly varies depending on the study area, soil condition, climatic condition, and even satellite sensors used (Albergel et al., 2008a).

4.3.3.2 Results after scaling

Similar to the results obtained from the ASCAT SWI-CDF, it is clear that the CDF matching method provides enhanced performance scores for most of the bias-correction groups with the exception of QM1 (Figure 4-7). The mean *r* values increase from 0.36 (AMSR2 SWI at 10 cm) to 0.39, 0.70, 0.60 and 0.68 for QM1, QM2, QM3 and QM4, respectively. The results obtained from QM1 are very similar to those derived from ASCAT, showing that the performance is apparently lower than the other groups. The QM2 based on a monthly duration shows the best performance among others: the RMSE ranges from 0.11 to 0.18, with the average value of 0.15; the *r* value is in the range 0.52-0.80, with the average value of 0.70.

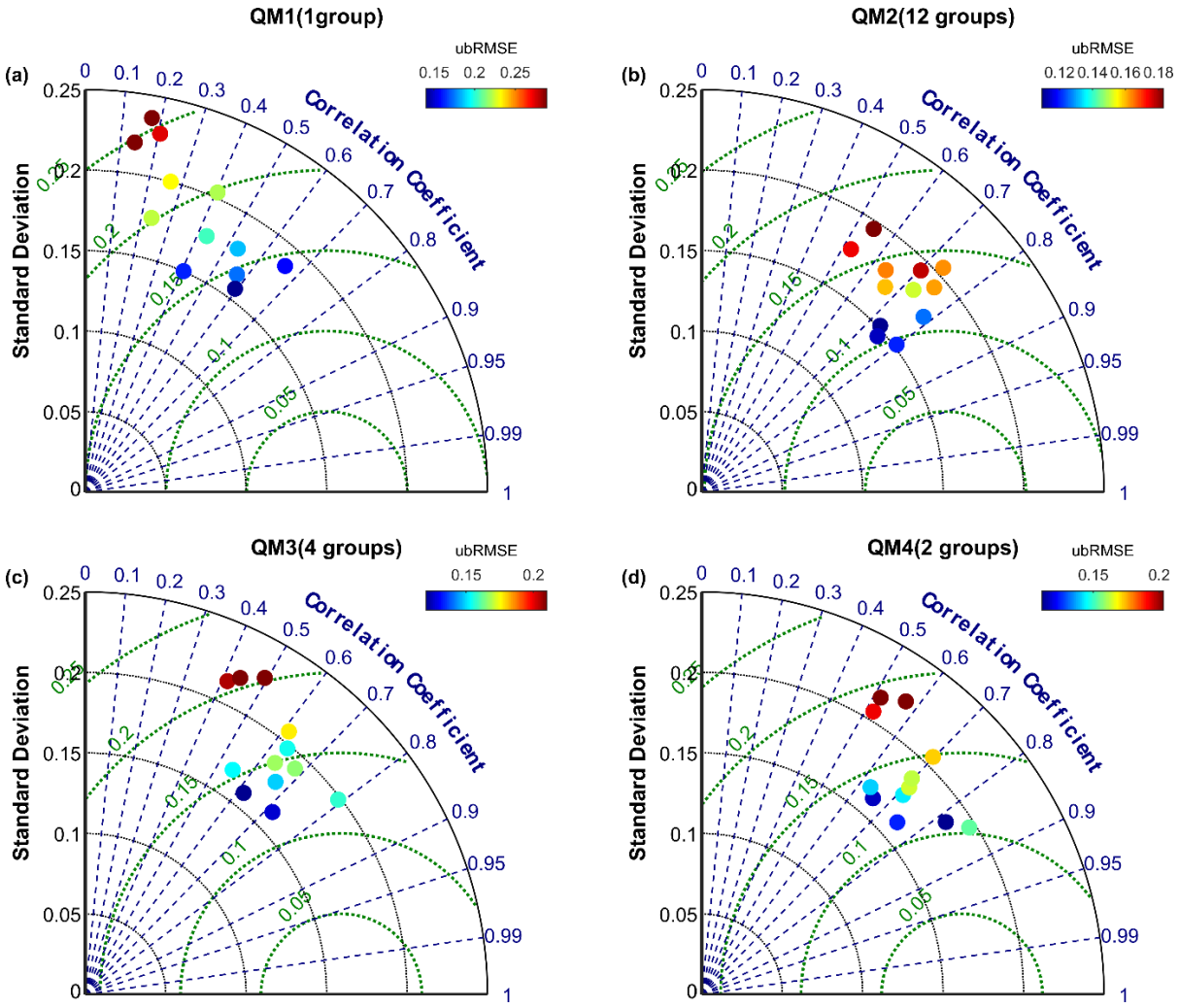


Figure 4-7 Taylor diagram representing the statistics between the in-situ observations measured at 10 cm depth and AMSR2 SWI-CDF at 12 sites.

Given that too many groups can cause serious overfitting issues, I subdivided datasets into two subsets and then validated the proposed CDF matching method through cross validation. As can be seen in Figure 4-8, it is evident that QM1 shows the worst performance in both calibration and validation periods. As for QM2 and QM3, significant different statistical scores are found between the calibration and validation periods resulting from overfitting issues. In contrast, QM4 shows a robust performance over both calibration and validation periods, thus confirming that the derived observation operator based on growing and non-growing seasons performs the best. These results are in line with the ASCAT.

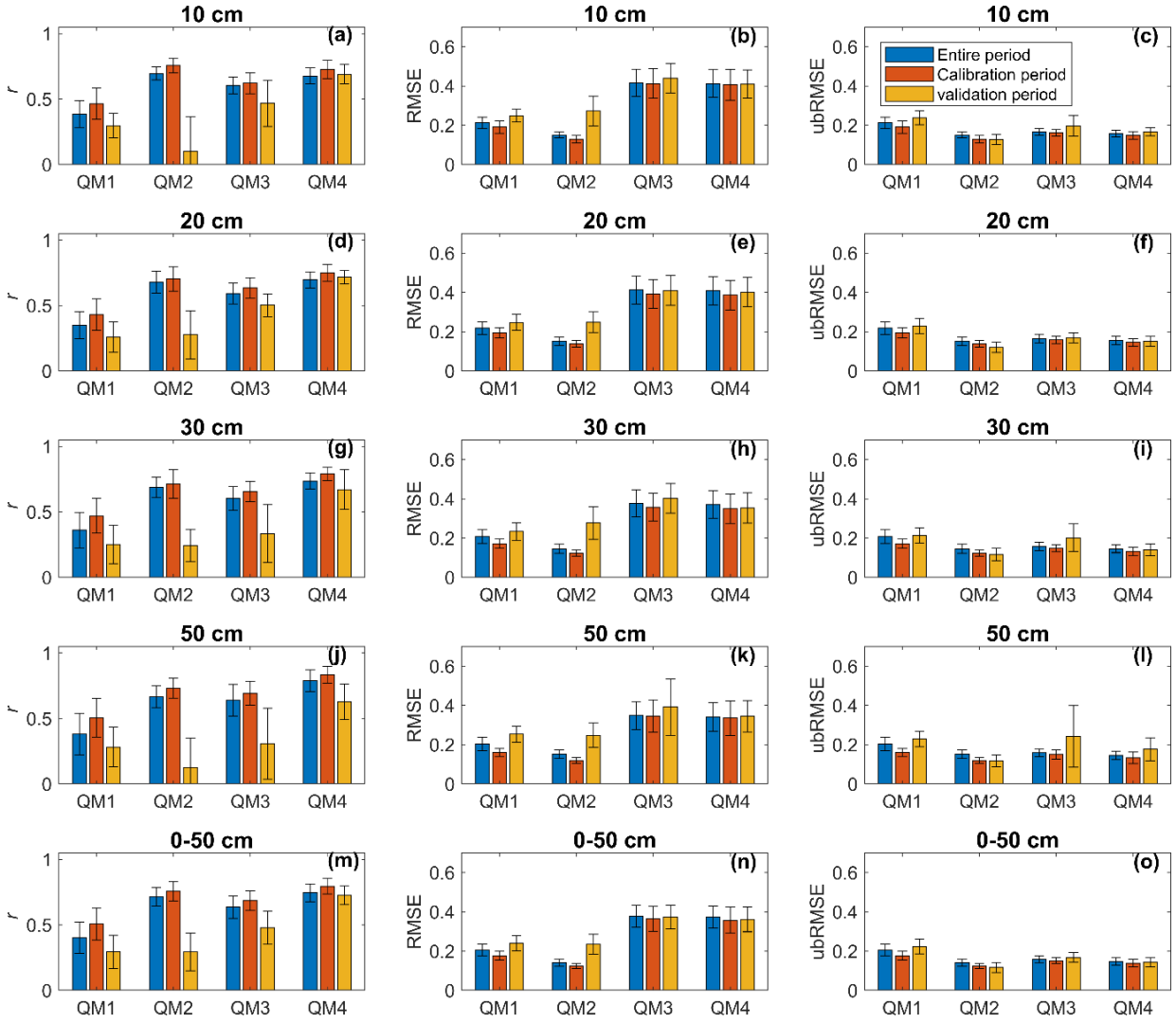


Figure 4-8 Statistics of the correlation coefficient (r), and RMSE. Here, the error bar indicates 95% confidence interval.

4.4 Conclusions

This chapter aims to assess active and passive microwave SM retrievals and further expand their applicability. I first estimated the accuracy of the original satellite SM retrievals in terms of their orbits as well as variation patterns. The key results obtained are summarised as follows.

(1) In terms of the ASCAT products, the descending overpass is more highly correlated with in-situ observations than the ascending overpass in this study area. Conversely, a slightly better

correlation is found in the ascending overpass for the AMSR2 although the differences are insignificant.

(2) As for the characteristic time T representing water travel time along the soil profile in the exponential filter method, it is found that the optimal T values generally increase with the depth of observed soil, which is in accordance with the model's underlying assumption. However, a smaller T value is obtained in the deeper soil layer at some observation sites, indicating that SM contents at the deeper layer tend to show rather larger temporal variability compared with that measured at the lower layer. Based on the results achieved in this chapter, it should be noted that although the determination of the optimal T value depends mainly on the soil depth, T value is also influenced by many uncertain factors, such as soil properties, length of data and climate conditions.

(3) Aside from the conventional bias correction approach that uses the whole record of the investigation period, I evaluated the performance of CDF matching method on a different temporal resolution basis to select an ideal combination: monthly (12 groups), seasonal (4 groups) and growing and non-growing (2 groups). The performance of each bias-correction group is then validated through a cross-validation procedure for the purpose of addressing overfitting issues. A bias-correction period of QM4 (2 groups; growing and non-growing) outperforms other temporal groups for both calibration and validation periods in South Korea.

Overall, the underlying features and limitations of satellite SM retrievals were investigated in depth, providing a meaningful perspective on the applicability of satellite SM products. The findings in this analysis further aided in addressing knowledge gaps in their varied performance in South Korea compared to different parts of the world. Moreover, successful attempts were made to overcome the shortcomings of the original satellite products, providing useful techniques concerning their practical applications; for instance, the exponential filter and bias correction method used in this chapter were widely employed for other hydrological applications throughout this thesis.

CHAPTER 5 [1]Spatial downscaling of soil moisture products

5.1 Motivation

In Chapter 4, considering the spatial mismatch and different measurement depths, a CDF matching method, as well as an exponential filter method, were employed to improve the applicability of satellite-derived SM. Aside from the abovementioned issue that satellite SM products are representative of a very shallow soil layer, another major challenge in using satellite SM data for practical applications is their coarse spatial resolution and uncertainties stemming from an inability to resolve sub-grid scale variability. To overcome this limitation, this chapter introduces a multivariate stochastic SM estimation approach based on GM-NHMM to spatially disaggregate the satellite SM data for multiple locations.

As illustrated in Section 2.4, many efforts have been made to achieve a higher spatial resolution for microwave SM data. However, the existing approaches all largely depend on a linear or nonlinear regression model to spatially downscale the satellite SM products without considering the uncertain factors affecting SM dynamics. The spatio-temporal dynamics of SM content result from complicated and mutually related processes of hydro-meteorological elements, such as subsurface flow, lateral flow, infiltration, precipitation, climate, and soil (Botter et al., 2007; Ridolfi et al., 2003). The influence of spatio-temporal variability in precipitation and temperature on the slow-varying behaviour of basin-scale SM can be better represented within a stochastic modelling framework (Botter et al., 2007). Recently, a stochastic downscaling technique, a nonstationary Markov model with a gamma (or exponential) distribution, has been widely used in both hydrology and meteorology (Khalil et al., 2010; Mehrotra and Sharma, 2005; Robertson et al., 2004).

---

[1]Moonhyuk Kwon, Hyun-Han Kwon and Dawei Han. "A spatial downscaling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian-mixture nonstationary hidden Markov model." *Journal of hydrology.* DOI: 10.1016/j.jhydrol.2017.12.015.

The stochastic downscaling approaches have been mainly used for rainfall simulation at multiple locations (Cioffi et al., 2017; Khalil et al., 2010; Kwon et al., 2009; Robertson et al., 2004). However, they have rarely been applied to SM data by means of a multivariate downscaling framework.

Given this background, this chapter aims to answer the following questions:

(1) Can daily SM sequences conditional on intraseasonal variability in climate be effectively clustered and discretized as a small set of states? In addition, can the identified states of daily SM and their transition probability be explicitly considered to better characterize SM dynamics?

(2) Is it desirable to use a nonstationary stochastic model that considers climate variables such as precipitation, temperature, and satellite-based SM products as predictors? Does a combination of climate variables and satellite-based SM better inform simulations?

(3) Can the proposed stochastic modelling framework be applied to simultaneously simulate the daily sequences of SM at multiple locations on a catchment scale?

In this chapter, I propose a GM-NHMM, which is primarily based on Hughes et al., (1999) and Yoo et al., (2015), to investigate those questions, with the intention of providing a practical tool for the estimation of daily SM on the catchment scale for use. In-situ SM observations at multiple stations are here used as a dependent variable, and both air temperature and rainfall, as well as the AMSR2 data, are considered as predictors. The proposed downscaling approach is applied to the Yongdam catchment in South Korea. The performance of the proposed downscaling scheme is then validated with 6 in-situ observations through a cross-validation procedure.

5.2 Study area and Datasets

In this chapter, I apply the spatial downscaling approach to the AMSR2 SM measurements for multiple stations. Figure 3-4 shows the study area and in-situ observation (i.e., SM and rainfall) sites. Air temperature was obtained from the Jangsu weather station operated by the KMA. The AMSR2 and ASCAT SM products were initially considered as potential predictors. As stated in Chapter 3, the AMSR2 SM data are in particular selected due to their relatively high temporal

resolution compared with the ASCAT (Figure 3-10). More importantly, the contribution of the ASCAT data in the proposed model was insignificant in comparison with that of the AMSR2 SM even though the original ASCAT data showed a slightly better agreement with in-situ observations as illustrated in Chapter 4. The proposed stochastic spatial downscaling scheme based on a nonhomogeneous hidden Markov chain model is then used to simulate SM data by mainly employing Gaussian mixture distributions over a predefined set of hidden states at multiple stations. The results point out that the overall accuracy of the original data does not ensure a better performance in the proposed model.

5.3 Methodology

5.3.1 Multivariate Gaussian-Mixture Nonstationary Hidden Markov Model

In this section, I briefly present only the relevant details of a multivariate hidden Markov model described elsewhere (Khalil et al., 2010; Kwon et al., 2011, 2009; Robertson et al., 2004; Yoo et al., 2015) and primarily based on Hughes et al. (1999). A schematic representation of the general procedure used to simulate the SM dynamics using GM-NHMM is illustrated in Figure 5-1.

A hidden Markov model (HMM) describes a process in which part of the system dynamics is hidden, and some other part of the system can be partially explained by other observations. The HMM uses a Markovian process and a set of stochastic functions to generate plausible sequences for a given time series based on stochastic sampling from probability distributions conditioned on different hidden states (Blunsom, 2004; Gharhramani, 2001).

Let SM_t be an M-dimensional vector of in-situ SM measurements corresponding to M-stations at time t. Let $SM_{1:T} = (SM_1, \dots, SM_T)$ denote a sequence of SM with length T. The sequence of observed SM measurements $SM_{1:T}$ is presumed to be governed by a Markov property with the corresponding sequence $S_{1:T} = (S_1, \dots, S_T)$ of a finite number of hidden states, taking on values k in {1, K}. A joint distribution of $SM_{1:T}$ and $S_{1:T}$ can be explicitly defined by taking the two conditional independence (CI) assumptions (Bishop, 2006; Smyth et al., 1997), as formulated below.

First, assume that the sequence of hidden states $S_{1:T}$ follows the stationary Markovian process that relies only on the values of the previous *k*-th order states. Obviously, the probability distribution for the current hidden state with a first-order model ($k=1$) can be represented as Eq. (5-1) (Rabiner, 1989).

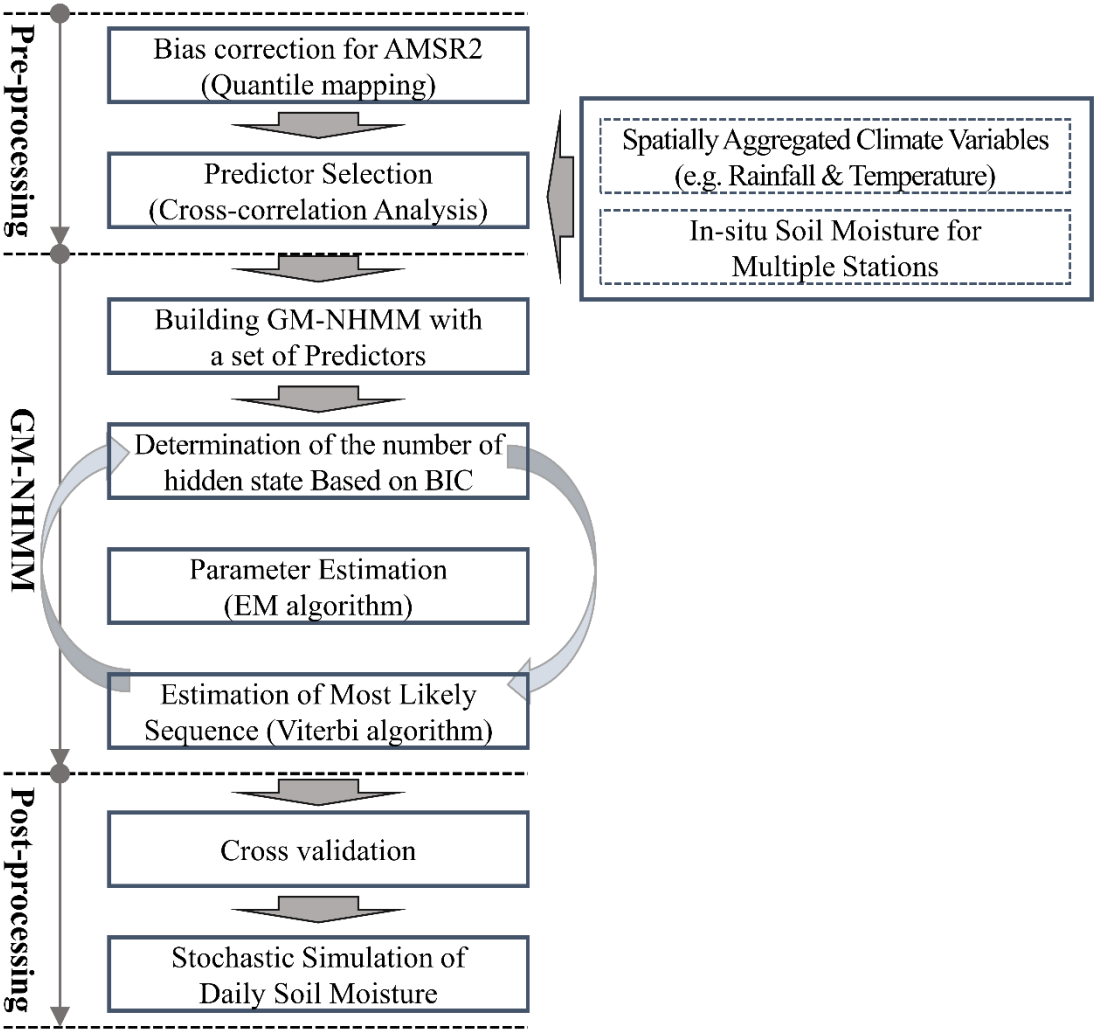


Figure 5-1 Schematic representation of the three-layer procedure of the proposed GM-NHMM approach for SM modelling.

$$p(S_1,...,S_T) = p(S_1)\prod_{t=2}^{T} p(S_t|S_{t-1}) \tag{5-1}$$

For a stationary HMM, $p(S_1)$ is the initial-state probability vector, and the state-transition probability matrix of a hidden state can be denoted as $p(S_t \mid S_{t-1}) = \{\gamma_{ij}\}$, $1 \le i, j \le K$.

Second, assume that individual in-situ observations SM_t are conditionally independent of all other variables in the model given the current state S_t (Robertson et al., 2006; Smyth et al., 1997).

$$p(\mathrm{SM}_{1:T}\,|\,\mathrm{S}_{1:T}) = \prod_{t=1}^{T} p(\mathrm{SM}_t\,|\,S_t)$$

(5-2)

The joint probability of the SM data $\mathbf{SM}_{1:T}$ and the hidden states can then be formulated as Eq. (5-3) (Kwon et al., 2011, 2009; Robertson et al., 2006).

$$p(\mathrm{SM}_{1:T},\mathrm{S}_{1:T}) = \left[\, p(S_1)\prod_{t=2}^{T} p(S_t\,|\,S_{t-1})\right]\left[\prod_{t=1}^{T} p(\mathrm{SM}_t\,|\,S_t)\right] \tag{5-3}$$

SM values, SM_t, at time t for M stations are assumed to be conditionally independent of one another given the hidden state S_t. Here, spatial dependencies across multiple stations are indirectly modelled by the hidden state variable, as described in Eq. (5-4). Note that a more advanced approach to modelling the spatial structure of SM_t across M sites could be of particular interest in situations with high spatial correlation. More specifically, the spatial coherence across stations is considered by assigning a state to each day, representing the spatial structure of SM (Kwon et al., 2011, 2009; Robertson et al., 2006).

$$p(\mathrm{SM}_t\,|\,S_t) = \prod_{m=1}^{M} p(\mathrm{SM}_t^m\,|\,S_t) \tag{5-4}$$

The probability density function for the emission distribution at an individual SM station $\mathbf{SM}_t^m$ is assumed to be approximated by a Gaussian mixture function of C components for non-zero SM, with $p_{i,m,c} \geq 0$ and $\sum_{c=1}^{C} p_{i,m,c} = 1$ for all $m = 1,\ldots,M$ and $i = 1,\ldots,K$, as follows:

$$p(\mathrm{SM}_t^m = r\,|\,S_t = i) = \sum_{c=1}^{C} p_{i,m,c} N(\mu_{i,m,c}, \sigma_{i,m,c}) \tag{5-5}$$

Here, μ and σ are the mean and variance of the Gaussian distribution, respectively, and the set of parameters associated with the transition matrix, the initial states, and the parameters of emission distribution are simultaneously estimated from the observed SM data using the expectation-maximization (EM) algorithm in an optimisation context. Gaussian mixture models are a statistical

tool for multimodal density estimation (Bilmes, 1998; Gauvain and Lee, 1994). Gaussian mixture models have been used for SM modelling (Ryu and Famiglietti, 2005; Verhoest et al., 2015; Vilasa et al., 2017), and have also been used extensively in hydrologic field (Carreau et al., 2009; Lakshmanan and Kain, 2010; Rings et al., 2012; Yoo et al., 2015). Unlike the HMM, the underlying assumption of the GM-NHMM is that SM is generated in a stochastic process that sequentially depends on a set of predictors represented by rainfall, temperature, and the satellite product. Specifically, NHMMs can be constructed by imposing a non-stationarity assumption on the probability distribution of the response variables, which in turn depends on observed independent variables (Hughes et al., 1999; Hughes and Guttorp, 1994; Kwon et al., 2011). This SM model can be substantially expanded by introducing a mixture model for SM content into the existing HMM. In this chapter, I use a mixture of Gaussians to describe SM at multiple stations in a stochastic framework to account for SM variability. Again, I use the expectation-maximisation (EM) algorithm to estimate the parameters, which is an iterative method for maximising the likelihood function (Dempster et al., 1977).

The concept of CI can be illustrated as edges in a directed acyclic graph of the GM-NHMM, as shown in Figure 5-2. Suppose $X_{1:T} = (X_1, \dots, X_T)$is a set of predictors representing SM, such as rainfall, temperature, and AMSR2 SM data. In a GM-NHMM, the state-transition matrix is assumed to be nonstationary, and therefore, the dynamic evolution of transition probability is a function of multivariate exogenous variables, $X_{1:T}$. The GM-NHMM is then written as Eq. (5-6) (Khalil et al., 2010; Kirshner, 2005; Kwon et al., 2011, 2009).

$$p(\mathrm{SM}_{1:T}, \mathrm{S}_{1:T} | \mathrm{X}_{1:T}) = \left[p(S_1 | \mathrm{X}_1) \prod_{t=2}^{T} p(S_t | S_{t-1}, \mathrm{X}_t) \right] \left[\prod_{t=1}^{T} p(\mathrm{SM}_t | S_t) \right] \tag{5-6}$$

In this study, I consider uniform priors, thus leading to the maximum likelihood approach to estimating a set of model parameters, $\arg\max_{\Theta} P(SM|X, \Theta)$. Again, note that the proposed model assumes that the observed SM sequences from different years are conditionally independent. Under the GM-NHMM, the log-likelihood function $LL(\Theta)$ of the observed SM data at multiple locations can be written as follows (Khalil et al., 2010):

$$LL(\Theta) = \ln p(\mathrm{SM}_{1:T} \mid \mathrm{X}_{1:T}, \Theta)$$
$$= \sum \ln \sum_{S_{1:T} \in [1,\ldots,K]^T} \left[p(S_1 \mid \mathrm{X}_1, \Theta) \prod_{t=2}^{T} p(S_t \mid S_{t-1}, \mathrm{X}_t, \Theta) \right] \left[\prod_{t=1}^{T} p(\mathrm{SM}_t \mid S_t, \Theta) \right]$$

(5-7)

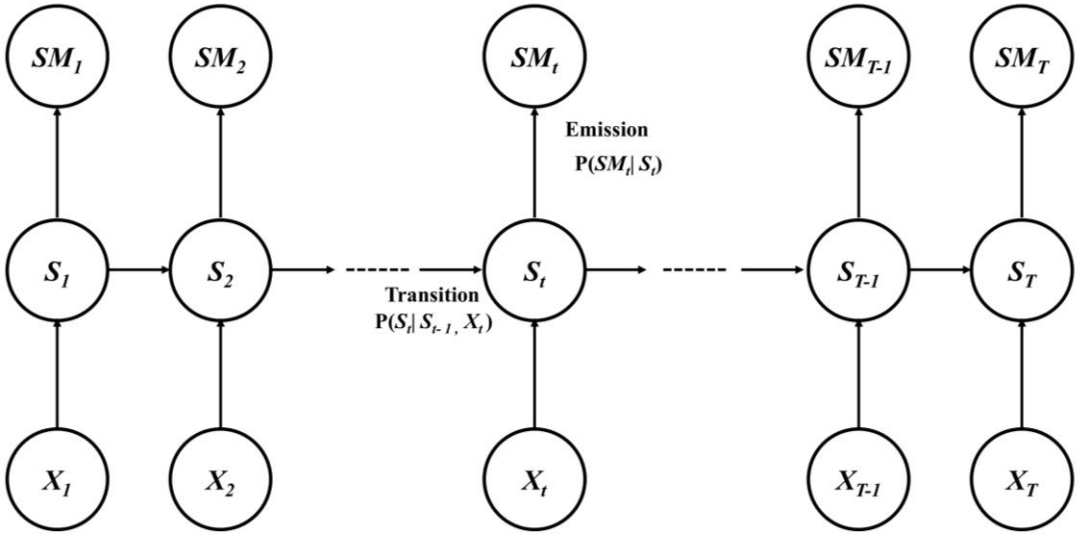


Figure 5-2 Graphical model representation of nonhomogeneous hidden Markov model. Here, *SM*, *S*, *X* indicate SM, hidden state and exogenous variable (i.e., rainfall, temperature, and AMSR2), respectively.

The parameter values cannot be obtained analytically, so I use the EM algorithm to estimate the value of the parameter vector Θ by maximising Eq. (5-7). The EM algorithm is an iterative method for maximising the likelihood function in a parameter space Θ. Finally, the state evolutions over time in Eq. (5-6) are simulated by a multinomial logistic regression as follows (Kirshner, 2005; Kwon et al., 2011):

$$p(S_t = \beta \mid S_{t-1} = \alpha, \mathrm{X}_t = \mathrm{x}) = \frac{\exp(\omega_{\alpha\beta} + \xi'_\beta \mathrm{x})}{\sum_{k=1}^{K} \exp(\omega_{\alpha\beta} + \xi'_k \mathrm{x})} \tag{5-8}$$

All the parameters ω are real and ξ is a vector in a multi-dimensional parameter space. Here, the prime denotes the transpose of the vector. Parameterization and prediction using NHMM are well documented in the statistical literature and, thus, need not be elaborated here. For a more detailed

description of the NHMM algorithm the reader is referred to Daniel and Martin, (2017), Gharhramani, (2001), Rabiner, (1989), and Robertson et al., (2003)

5.3.2 Ordinary Linear Regression (OLR)

As a comparison to the GM-NHMM, I applied a linear regression model with the same input variables used in the GM-NHMM to downscale the AMSR2 SM product for each station *m*. Here, each parameter (β) is obtained from the least squares method. The linear combination of predictors for estimating SM can be written as follows:

$$SM_t^m = (\beta_0^m + \beta_1^m \times R_t + \beta_2^m \times Tp_t + \beta_3^m \times ST_t) \quad (5\text{-}9)$$

where *SM*, *R*, and Tp are in-situ SM, rainfall, and temperature data, respectively, and *ST* is 10km AMSR2 SM data. Again note that predictor variables used here are averaged over the entire region.

5.4 Results and discussion

5.4.1 CDF matching for bias correction

Prior to disaggregating the AMSR2 SM products, this chapter also employs the CDF matching method with intent to remove the systematic bias from the satellite SM data for subsequent use. In this study, t location-scale (eq. (5-10)) and gamma (eq. (5-11)) distributions were selected to fit the AMSR2 and in-situ SM data, respectively, based on the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978), respectively, as summarised in Table 5-1.

Table 5-1 BIC and AIC scores with respect to distribution models.

In-situ			AMSR2		
Distribution	**BIC**	**AIC**	**Distribution**	**BIC**	**AIC**
Gamma	44,677	44,663	t-location scale	31,445	31,425
Log-logistic	45,051	45,037	Log-logistic	32,316	32,303
Normal	45,128	45,114	Gamma	36,550	36,536
t-location scale	45,137	45,116	Weibull	38,680	38,666
Weibull	45,259	45,246	Normal	43,660	43,646

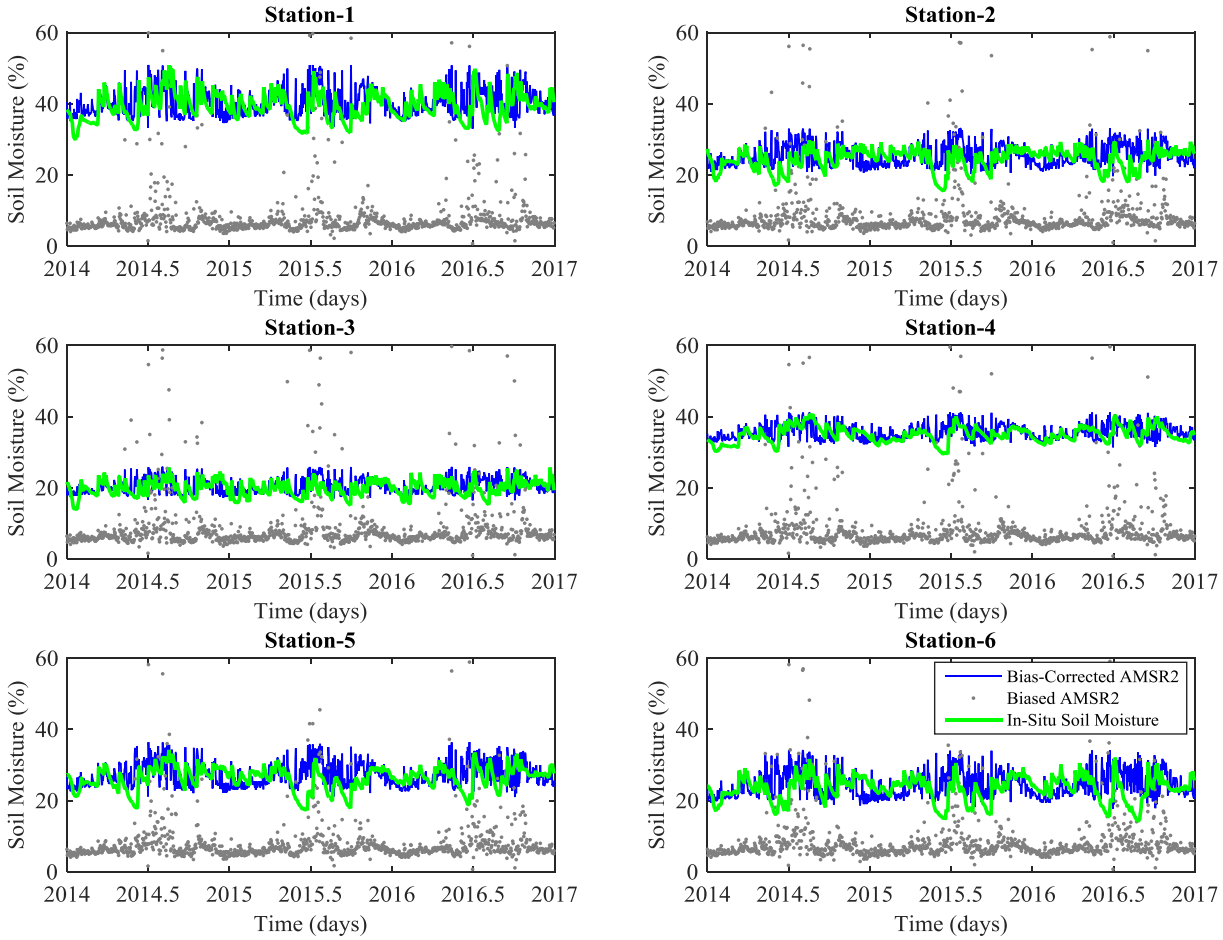


Figure 5-3 Bias-uncorrected and bias-corrected AMSR2 SM time series data with in-situ observations during the study period, 2014–2016.

As shown in Figure 5-3, the bias-corrected AMSR2 SM data exhibit enhanced variability and match well with the in-situ observations. I used these bias-corrected AMSR2 SM products for this study.

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+(\frac{x-\mu}{\sigma})^2}{\nu}\right]^{-\left(\frac{\nu+1}{2}\right)} \tag{5-10}$$

$$f(y) = \frac{y^{\theta-1}e^{-\tau y}}{\tau^{\theta}\Gamma(\theta)} \tag{5-11}$$

where μ, σ, and ν are the location, scale, and shape parameters of the t location-scale distribution, respectively, and $\Gamma(\bullet)$ is the gamma function. θ and τ are the shape and scale parameters of the gamma distribution, respectively.

5.4.2 Predictor selection

It is important to identify a suitable set of predictors that consistently influences the response variables. However, in a regression model, using several predictors can cause serious overfitting, which results in unrealistic predictions (Khalil et al., 2010). For a parsimonious model, I consider only three predictors, daily rainfall, air temperature, and AMSR2 data, and I initially evaluate the cross-correlations for all lagged orders. The correlations are statistically significant and strongly persistent, as illustrated in Figure 5-4. Note that here the values are averaged over the entire watershed for a representation. The lag-1 correlation is high for daily rainfall, and the correlations appear to be consistent with the lag in the temperature and AMSR2 data. Therefore, I retained a set of 1 day time-lagged values for the three predictors to simulate SM content in the proposed GM-NHMM.

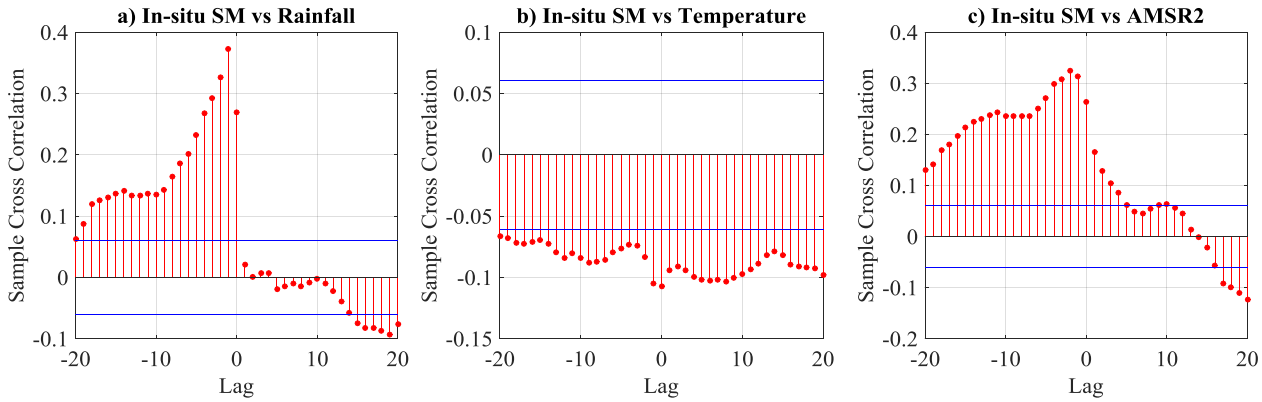


Figure 5-4 Sample cross correlation between the in-situ SM and a set of predictors: a) rainfall, b) temperature, and c) AMSR2 SM data. All values are averaged over the entire watershed.

5.4.3 Stochastic modelling of SM using GM-NHMM

The performance of the GM-NHMM is greatly influenced by the number of hidden states used to represent an unobserved SM state. In this study, I estimated the number of hidden states by recursively maximising the log-likelihood (or minimizing the BIC) in the context of optimisation.

The maximized log-likelihoods for each state are shown in Figure 5-5, together with the minimized BIC. As shown in Figure 5-5(a), the log-likelihoods gradually increase with the number of hidden states, but I could not clearly identify an inflexion point on the curve to determine the optimal number of hidden states. On the other hand, the BIC decreases rapidly at 4 states, and the degree of reduction beyond 6 hidden states is negligible. Therefore, I used 6 hidden states to build a stochastic SM model at multiple locations. For the selected 6 hidden states, the most likely temporal sequences can be efficiently determined using the Viterbi algorithm (Viterbi, 1967), which calculates the probability that a hidden state will occur as well as the probability that it will transition to another state at a certain date. The estimated temporal sequences of observed SM are illustrated in Figure 5-6, and considerable inter-annual and intraseasonal variability are clearly identified. The Viterbi analysis is a useful tool not only to capture intra- and inter-annual variability but also to quantify its intensity. More specifically, changes in the intra-annual sequence of observed SM states are shown along a horizontal line, and inter-annual variability is represented by a vertical line.

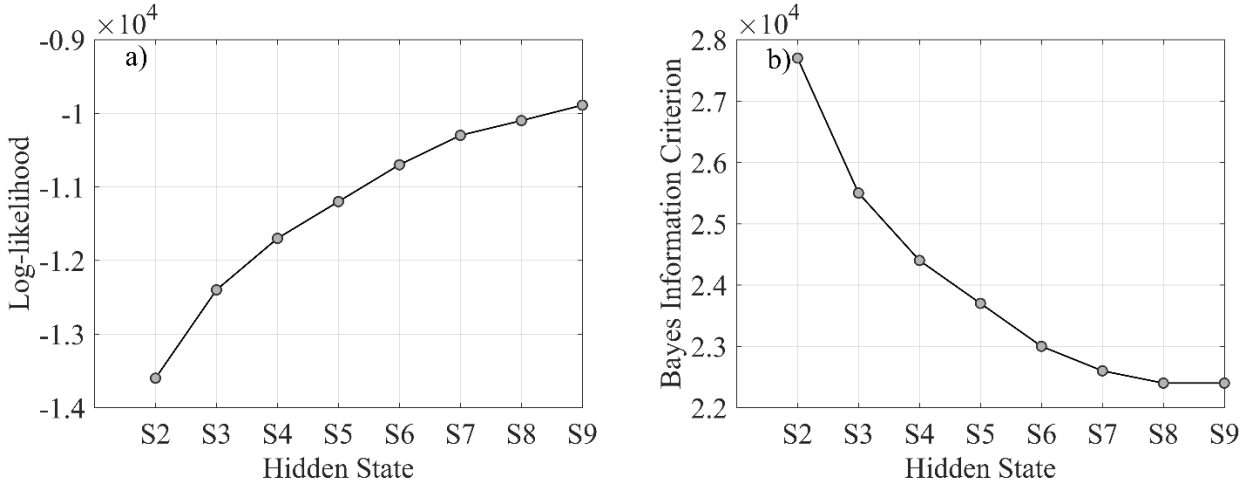


Figure 5-5 Log-likelihood and BIC values in terms of hidden states.

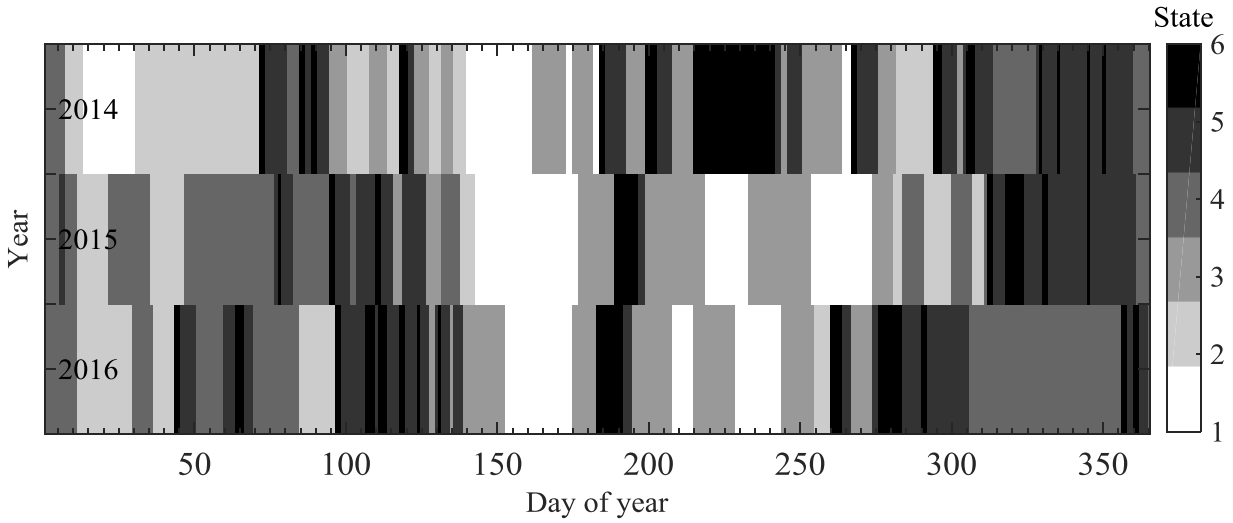
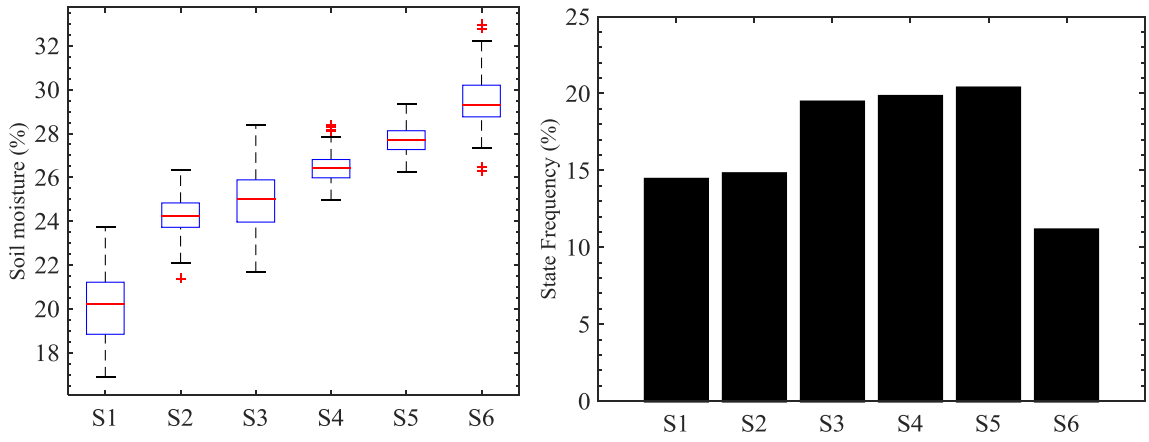


Figure 5-6 Estimated hidden state sequence for a 3-year period (2014–2016).

The degree of soil wetness and the frequencies associated with hidden states are presented in Figure 5-7. Figure 5-7(a) shows boxplots representing station-averaged SM data corresponding to each state in 2014–2016. Clearly, the lower states are closely related to drier soil conditions, and vice versa. Moreover, the median SM value increases largely as a function of the number of states (i.e., from 21% (state 1) to 29.3 % (state 6)). The percentage of days falling into the 6 hidden states for SM data across 6 stations are 14.4, 14.8, 19.5, 19.8, 20.3, and 11.1 %. States 3–5 occur dominantly during the entire period, accounting for 59.6 %, whereas state 6, representing the wettest soil condition, has the lowest frequency, as shown in Figure 5-7(b). The estimated transition probabilities of the NHMM are shown in Table 5-2. Note that the state-transition in the GM-NHMM is assumed to be nonstationary and informed by exogenous variables, such as rainfall and temperature. As expected, the self-transition probability (more likely to stay in the current state than to transition to a new state) is noticeably high, with state 1 being the most persistent (0.93) and state 6 being the least persistent (0.70).



(a) SM per state (b) Frequency of SM per state

Figure 5-7 Estimated distribution and frequency of SM in each state.

Table 5-2 Transition probability matrix of 6 hidden states for SM at 6 stations in the Yongdam watershed.

	Station 1	**Station 2**	**Station 3**	**Station 4**	**Station 5**	**Station 6**
Station 1	0.93	0.01	0.05	0.00	0.00	0.01
Station 2	0.02	0.90	0.01	0.04	0.00	0.04
Station 3	0.04	0.03	0.88	0.00	0.03	0.02
Station 4	0.00	0.04	0.00	0.92	0.02	0.02
Station 5	0.00	0.00	0.07	0.05	0.79	0.09
Station 6	0.00	0.00	0.00	0.00	0.30	0.70

The temporal patterns of the simulated SM and the in-situ observations at 6 stations are illustrated in Figure 5-8. To verify the potential of the model to reproduce the variability observed in the SM data, I conducted 100 simulations. The results show a fairly good agreement with the in-situ observations. Here, the proposed GM-NHMM is illustrated across the entire period (2014–2016), along with the OLR model, in Figure 5-9. The GM-NHMM comprises the vector of observed SM data from 6 stations (as dependent variables) given a vector of observed covariates (as independent variables). For comparison, I built an OLR model for each station using the ordinary least square method for the best-fit model of SM data. Summary statistics for the comparison between the GM-

NHMM and OLR are presented in Table 5-3, and the GM-NHMM outperforms the OLR model. More specifically, the SM data simulated through the GM-NHMM agree well with the in-situ observations, with correlation coefficients (*r*) ranging from 0.73 to 0.81 (mean: 0.78), and a root mean square error (RMSE) ranging from 1.47 % to 2.62 % (mean: 2.06 %), whereas the OLR has much lower performance (mean *r*: 0.49 and mean RMSE: 2.58 %).

Table 5-3 Comparison between in-situ and simulated SM. Here, BC AMSR2 and GM-NHMM refer to goodness of fit obtained from biased corrected AMSR and Gaussian-mixture nonstationary hidden Markov model, while OLR represents the results obtained from ordinary linear regression model during the study period (2014-2016).

Site	BC AMSR2		GM-NHMM		OLR	
	r	**RMSE (%)**	***r***	**RMSE (%)**	**r**	**RMSE (%)**
Station 1	0.34	4.55	0.79	2.62	0.49	3.36
Station 2	0.10	4.07	0.78	2.02	0.55	2.42
Station 3	0.31	2.55	0.73	1.52	0.49	1.83
Station 4	0.38	2.54	0.81	1.47	0.54	1.86
Station 5	0.17	4.34	0.79	2.22	0.41	2.95
Station 6	0.10	4.93	0.79	2.50	0.48	3.06
Average	0.23	3.83	0.78	2.06	0.49	2.58

To further ensure that the proposed modelling scheme can predict SM, I subdivided the SM data into different groups and then validated the proposed GM-NHMM using a cross-validation scheme. I partitioned a sample of SM data into three different subsets corresponding to the year of interest, trained the model on one subset, and then validated the model with the remaining data. In other words, a set of parameters for the GM-NHMM is estimated in the training period, and the identified parameters are then used to simulate SM for the validation. I performed 100 simulations for each cross-validation partition for both the training and validation periods. As a representative case, the simulated SM values for 6 stations are compared with the values observed at those stations for the training period (2014–2015) and the testing period (2016) in Figure 5-10. The SM data are reasonably well reproduced by the proposed GM-NHMM for both the training and testing phases. The results of the cross-validation using the GM-NHMM for the different partitions are summarised in Table 5-4.

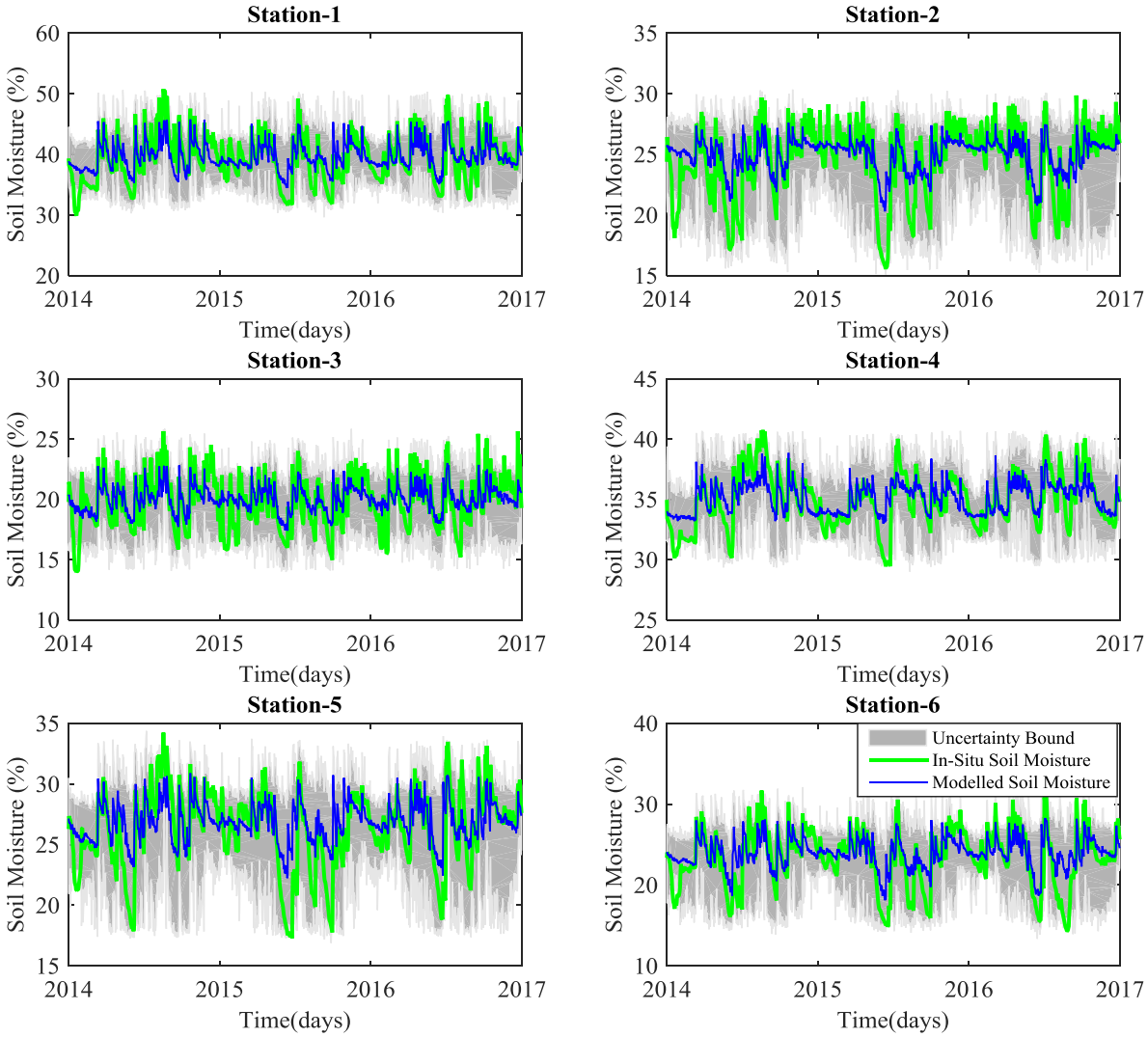


Figure 5-8 Comparison of time series data between the in-situ and GM-NHMM-simulated SM data for 2014–2016: the green line indicates the in-situ observations, and the blue line represents the median of 100 simulations. The shaded area represents the uncertainty bound of simulations (between 2.5% and 97.5%).

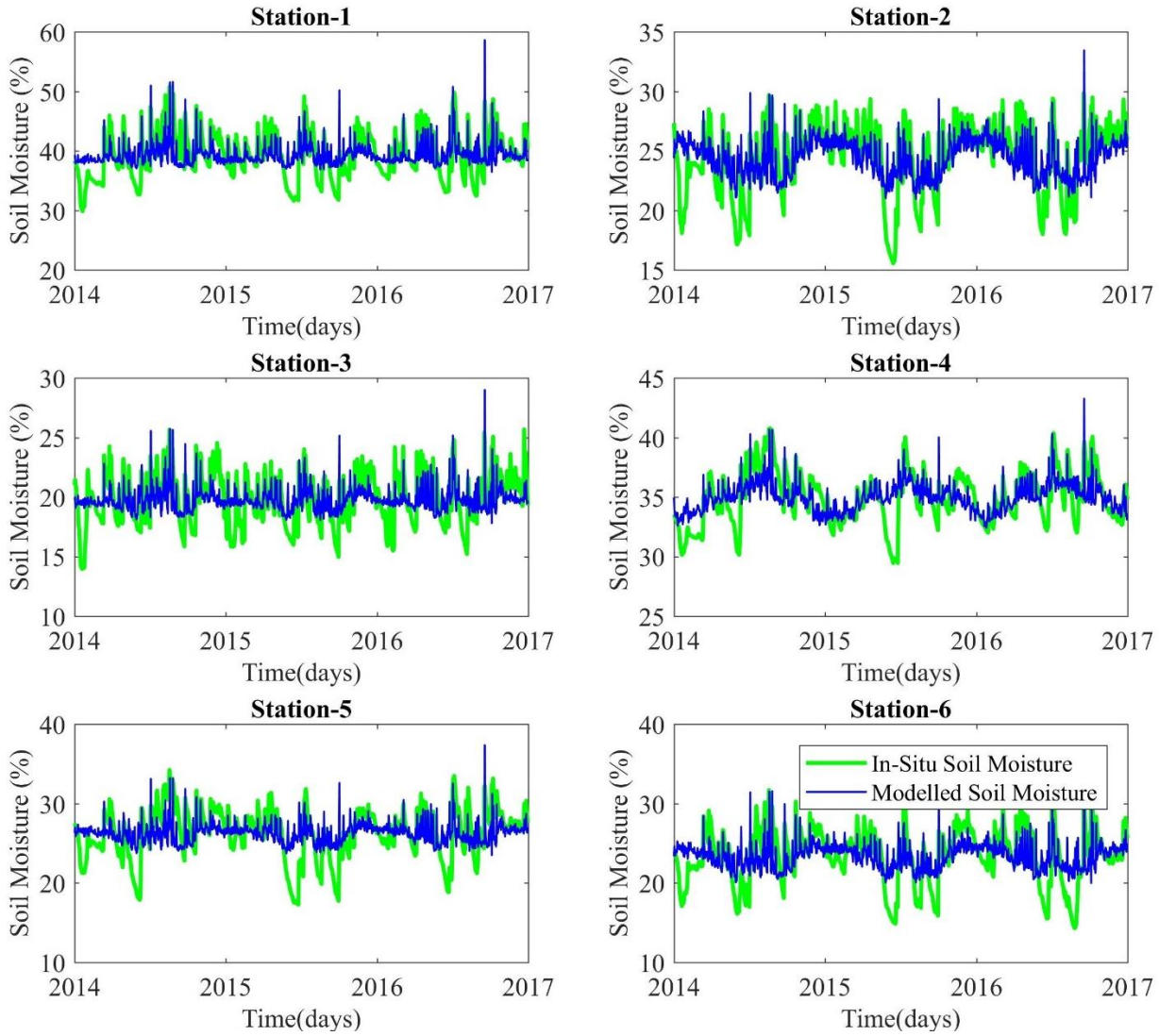


Figure 5-9 Comparison of time series data between the in-situ and OLR-simulated SM products for 2014–2016: the green line indicates in-situ observations, and the blue line represents OLR-simulated SM.

I considered three goodness-of-fit measures, correlation coefficient (*r*), RMSE, and bias, in evaluating the models. During the training periods, the 6-station averaged correlation coefficient values range from 0.72 to 0.80, whereas during the validation period, the *r* values show slightly lower correlations than during the training period. However, the GM-NHMM can clearly generate the intraseasonal sequence of daily SM fairly well, and other measures also show reasonable performance at multiple locations, leading to higher correlations with the observed SM data. The RMSE and bias values are also generally better for the training period than the validation period.

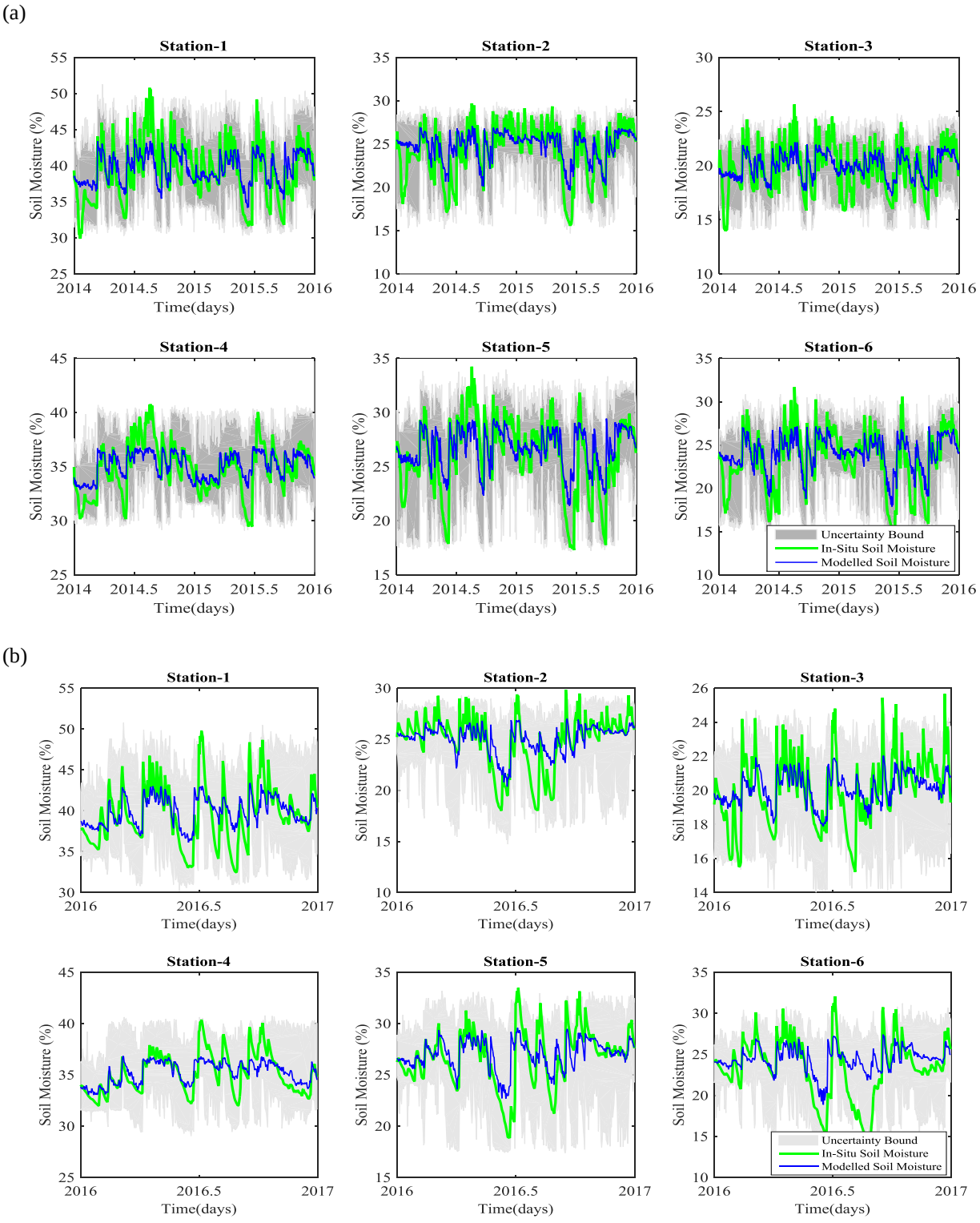


Figure 5-10 Comparisons between the sequences of simulated SM and that observed at multiple locations in the Yongdam watershed for a) the training period (2014–2015) and b) the validation period (2016).

Table 5-4 Comparison between in-situ and simulated SM.

Site	Training (2014–2015)			Validation (2016)			Training (2015–2016)			Validation (2014)			Training (2014, 2016)			Validation (2015)		
	r	**RMSE (%)**	**Bias**	***r***	**RMSE (%)**	**Bias**	***r***	**RMSE (%)**	**Bias**	***r***	**RMSE (%)**	**Bias**	***r***	**RMSE (%)**	**Bias**	***r***	**RMSE (%)**	**Bias**
Station 1	0.79	2.69	0.43	0.80	2.47	0.32	0.83	2.14	0.32	0.62	3.34	0.26	0.77	2.79	0.72	0.68	3.14	1.53
Station 2	0.79	2.10	0.66	0.75	1.85	0.19	0.86	1.65	0.37	0.63	2.57	1.36	0.73	2.25	0.83	0.86	2.26	0.69
Station 3	0.76	1.49	0.28	0.67	1.57	0.10	0.75	1.45	0.25	0.69	1.65	0.00	0.68	1.70	0.44	0.74	1.80	1.02
Station 4	0.80	1.57	0.24	0.83	1.24	0.06	0.73	1.44	0.20	0.74	1.86	0.06	0.76	1.57	0.35	0.59	1.91	0.99
Station 5	0.79	2.37	0.67	0.78	1.89	0.23	0.76	2.18	0.48	0.60	2.63	0.19	0.65	2.54	0.87	0.66	3.47	2.11
Station 6	0.83	2.23	0.68	0.73	2.96	1.04	0.88	1.88	0.41	0.68	2.43	0.30	0.71	2.80	1.01	0.86	2.60	1.27
Average	0.79	2.08	0.49	0.76	2.00	0.33	0.80	1.79	0.34	0.66	2.41	0.36	0.72	2.28	0.70	0.73	2.53	1.27

For a multisite SM simulator, it is of particular importance to correctly reproduce the spatial coherence of daily SM across multiple stations. Therefore, I estimated the spatial correlations of the sequence of daily SM and compared them with the observed values. As shown in Figure 5-11, the spatial correlations across the stations are reasonably well reproduced by the proposed GM-NHMM model.

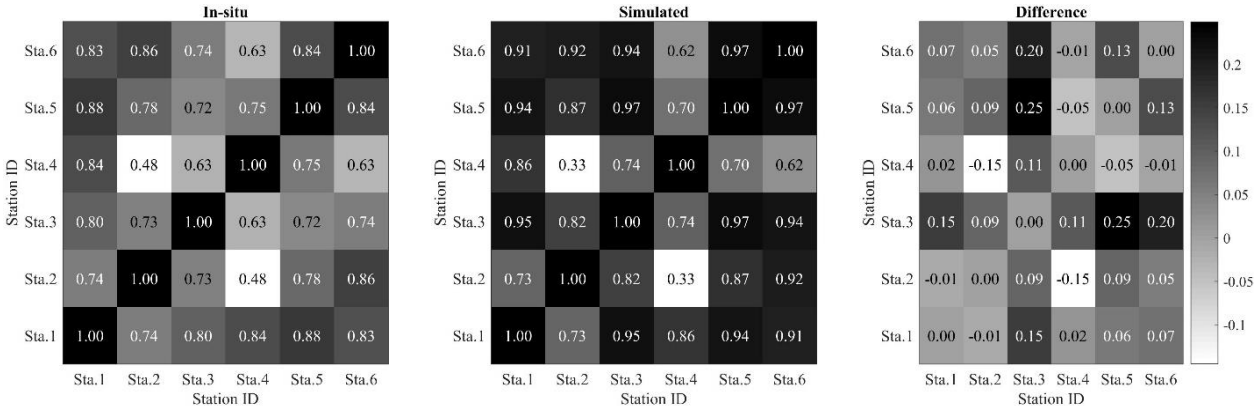


Figure 5-11 Comparison of the spatial correlation matrices between the observations and simulations of daily SM sequences across 6 stations.

Table 5-5 Comparison of *r* values with respect to different combinations of predictors.

Sta. No	Modelling	Cross Validation					
	Entire period (2014–2016)	Training (2014–2015)	Validation (2016)	Training (2015–2016)	Validation (2014)	Training (2014, 2016)	Validation (2015)
(Case 1) Predictors: Rainfall, Temperature							
Station 1	0.75	0.76	0.72	0.76	0.64	0.71	0.59
Station 2	0.73	0.74	0.70	0.84	0.60	0.56	0.74
Station 3	0.63	0.70	0.48	0.69	0.66	0.52	0.65
Station 4	0.78	0.78	0.79	0.70	0.71	0.81	0.66
Station 5	0.73	0.75	0.68	0.70	0.54	0.64	0.42
Station 6	0.75	0.77	0.72	0.87	0.60	0.55	0.66
Average	0.73	0.75	0.68	0.76	0.63	0.63	0.62
(Case 2) Predictor: Rainfall							
Station 1	0.78	0.78	0.79	0.72	0.81	0.80	0.70
Station 2	0.39	0.45	0.22	0.23	0.51	0.38	0.47
Station 3	0.62	0.66	0.53	0.57	0.67	0.56	0.62
Station 4	0.81	0.80	0.84	0.79	0.83	0.84	0.70
Station 5	0.62	0.61	0.64	0.49	0.75	0.67	0.53
Station 6	0.57	0.61	0.50	0.49	0.67	0.59	0.63
Average	0.63	0.65	0.58	0.55	0.71	0.64	0.61

Table 5-5 shows the results of applying the GM-NHMM with different combinations of predictors to examine the contribution of the AMSR2 SM data to the proposed model. The use of rainfall and temperature without the AMSR2 data (case-1) led to a slightly lower correlation coefficient of 0.73, compared to the results obtained with all three predictors shown in Table 5-5. On the other hand, there was no significant change in the correlation coefficient of 0.63 when I used rainfall alone as a predictor (case-2). Furthermore, I found a similar trend in our cross-validation analysis. Therefore, the 1 day time-lagged rainfall data might be the main factor in properly reproducing SM dynamics. Nonetheless, combining rainfall with temperature and AMSR2 still yielded the highest correlation with the in-situ observations.

5.5 Conclusions

In this chapter, I have presented a stochastic SM estimation model based on a GM-NHMM to spatially disaggregate AMSR2 SM data at multiple locations in the context of downscaling. Given the close relationship with SM, both rainfall and air temperature were considered as potential predictors in the proposed stochastic downscaling model. 1 day time-lagged values for the three predictors (i.e., Rainfall, temperature and AMSR2) were also used in the proposed GM-NHMM model. Before applying the proposed downscaling scheme, the quantile mapping approach was employed to reduce the systematic bias in the AMSR2 SM products, and those bias-corrected AMSR2 SM products were exploited for subsequent analyses. In GM-NHMM terms, I formulated a six-state model with three predictors representing an unobserved SM state based on the BIC. The temporal sequences of unobserved hidden states and the dynamic evolution of transition probability were estimated by the Viterbi algorithm. Consequently, the proposed GM-NHMM was applied to simulate fine-resolution SM products in a multivariate framework. I compared the results with in-situ observations from the Yongdam catchment in South Korea. The key results obtained in this chapter are summarised as follows.

(1) The estimated small set of hidden states that most likely corresponds to localized SM dynamics is effectively captured and accounts for a certain fraction of the SM process, which improves understanding of the intraseasonal and inter-annual variability of SM dynamics. Based on the identified state transition-probability matrix, self-transitions are more significant than the

probability of transitioning to other states, indicating that the states seem to be persistent over time due to the slow-varying behaviour of basin-scale SM (Botter et al., 2007).

(2) Given the relatively short length of the in-situ SM time series data, I considered a cross-validation performance assessment of the simulations. The results presented in this chapter illustrate the potential of a stochastic model with a climate-predictor-based forecast. However, the relatively small improvement in forecast skill that the AMSR2 SM products offer in the model suggests that the AMSR2 data might not sufficiently reflect the regional or seasonal characteristics of this study area.

(3) I compared the efficiency of the proposed model with that of an ordinary regression model using the same predictors. The mean correlation coefficient for the GM-NHMM obtained by averaging over all the stations is about 0.78, which is significantly greater than that of the OLR, about 0.49. The proposed model also yields a noticeable reduction in RMSE.

The main contributions in this chapter are insights into the SM process and its potential predictability, leading to the way for more applications in hydrologic studies. The findings and approaches used in this chapter are expected to offer useful guidelines for acquiring fine-resolution SM product. Moreover, given that most global-scale SM datasets are provided to the public domain with coarse spatial scale (10-40 km), the downscaling scheme can be effectively applied to other SM datasets including the land surface model (e.g., the ERA-Interim and GLDAS SM product) for regional applications.

CHAPTER 6 [1]Incorporating soil moisture state into a hydrological model

6.1 Motivation

Previous chapters (Chapter 4 and Chapter 5) focused on overcoming the inherent limitations of the use of satellite SM products. This is, the scaling/filtering method was employed to mitigate mismatch in observation depths and associated biases in Chapter 4, while the critical issue arising from a coarse spatial resolution was addressed by introducing a stochastic downscale scheme in Chapter 5. In this chapter, considering the importance of identifying spatio-temporal variation in SM which aids in a better understanding the hydrological behaviour, the degree of contribution of SM and its particular role in RR simulation is explored by incorporating satellite SM into the proposed RR modelling framework.

The hydrological model used for simulating the RR process is an imperative tool for water resource planning and management. To date, despite considerable efforts and improvements to establish an effective RR model for a better understanding of catchment response to rainfall event, there still remains the question whether the improvement meets scientific and practical demands (Orth et al., 2015). Clearly, apart from the model formulation, SM appears as a key variable in understanding the interaction between surface rainfall and water loss, which can be mainly attributed to the corresponding hydrological elements such as evapotranspiration, infiltration and percolation loss (Dorigo et al., 2017; Massari et al., 2015). The performance of runoff modelling is therefore significantly dependent on how accurately the RR model can capture the spatio-temporal variation of SM conditions (Brocca et al., 2010; Loizu et al., 2018). Despite the importance of identifying a spatio-temporal variation in SM, which in turn lead to a better understanding of the hydrological behaviour, its use in the hydrological model for input is limited by the lack of an in-situ observation network in many parts of the world including South Korea.

[1]Moonhyuk Kwon, Hyun-Han Kwon and Dawei Han. "A hybrid Approach Combining Conceptual Hydrological Model, Support Vector Machine and Remote Sensing Data for Rainfall-Runoff Modeling." *Water Resources Research.* (under review).

As stated in Section 2.2, over the past decades, satellite-based SM products have become available with increased accuracy and frequency (Loizu et al., 2018) and their hydrological applications have substantially expanded and used for an alternative data source to explore the various issues and challenges involved in different aspects of hydro-meteorological problems such as numerical weather prediction (Dharssi et al., 2011), rainfall estimation (Brocca et al., 2014; Ciabatta et al., 2018) and drought monitoring (Enenkel et al., 2016). In addition to the aforementioned applications, satellite SM retrievals have been integrated into either conceptual (Brocca et al., 2012; Massari et al., 2015) or physically-based RR models (Lievens et al., 2015; Loizu et al., 2018; Young et al., 2017) using data assimilation techniques. However, the benefits of the use of SM in the RR model varies greatly depending on the study, providing either no (or limited) improvement in the accuracy of runoff simulation or vice versa. More specifically, Lievens et al., (2015) found that SM made a limited contribution to runoff prediction, while Brocca et al. (2012) and Massari et al. (2015) achieved significantly improved overall response by combining satellite SM with conventional RR modelling framework. Conversely, no significant improvement was found by Han et al. (2012).

In this perspective, this chapter investigates a hybrid RR model that can make use of the intermediate state variables obtained from a conceptual model (i.e. the Tank model) within a least squares support vector machine (LSSVM; an upgraded version of the support vector machine) based regression framework. Here, I specifically used the LSSVM since it has a computational advantage over conventional support vector machines by converting a quadratic formulation into a set of linear equations (Kisi and Parmar, 2016; Okkan and Serbes, 2012). The fundamental hypothesis behind the proposed hybrid approach is that the combined use of different models may intuitively have more relevance to capture complex features of the RR processes, thus providing a better representation of the runoff process than the individual use. I first build the Tank model and explore the performance of the model in terms of simulating daily runoff discharge in the Yongdam catchment, South Korea. Here, I assume that the water depths at each tank are the intermediate state variables to be considered as the proxy measure of the SM. The role of the obtained water depths at each tank and their different combinations in the LSSVM based RR modelling scheme are then explored. In addition, the potential use of remotely sensed SM products and its combined use with the state variables in hydrological modelling are further examined in the context of the hybrid simulation.

Given this background, this study attempts to address the following questions:

(1) How much information on SM can implicitly be obtained from a conceptual RR model? How much are the captured intermediate variables correlated with in-situ SM?

(2) Can machine learning technique effectively describe RR relationship with a small set of predictors in the context of data-driven learning?

(3) Can the use of the obtained intermediate variables be more effective in RR modelling within a machine learning based regression framework, particularly for low flow simulation in the context of a hybrid model?

6.2 Study area and datasets

As mentioned in Chapter 3.1, the Yongdam catchment has a drainage area of 930 km^2 and primary land uses consist of forest, agriculture and urban, covering 70.1 %, 19.2 % and 2.6 % of the catchment area (Table 3-1), with an annual rainfall of 1,299 mm and runoff depth of 680 mm during the period 2007-2016. Hydrologic data such as runoff, rainfall and SM records are measured by the K-water. There are six hydrologic stations where precipitation has been measured since 2001, whereas SM has been recorded since 2014 using the TDR (Topp et al., 1980). Study area and the location of stations used in this chapter are presented in Figure 3-4. Here, the areal rainfall is defined by Thiessen Polygons, whereas climate data (i.e., air temperature, relative humidity, wind speed, and sunshine hour), required for estimating reference evapotranspiration, are obtained from Jangsu weather station operated by the KMA. In addition to meteorological data, the ESA CCI SM product (v04.2) is used in this chapter. As mentioned in Section 3.2, for the study period 2007-2016, the active-passive combined SM products with a spatial resolution of 0.25° are chosen due to their higher temporal resolution and better accuracy compared to either the active or the passive retrievals (Liu et al., 2011).

6.3 Methodology

6.3.1 Tank model

The tank model is classified as a deterministic, conceptual and continuous model. Similar to other conceptual models that in general consist of a series of interconnected subsystems (i.e., storages), the tank model used in this study is composed of three vertically interconnected tanks (i.e. 3-Tank model), as illustrated in Figure 6-1. The structure of the tank model such as the number of tanks and their associated side outlets in each tank can be effectively formulated in terms of physical catchment attributes and climatological conditions for a particular purpose. For instance, a model with two tanks was proposed to evaluate the RR relationship for a paddy field (Basri, 2013), while Paik et al. (2005) used a 4-Tank model for a better representation of deep percolation process in the forest region.

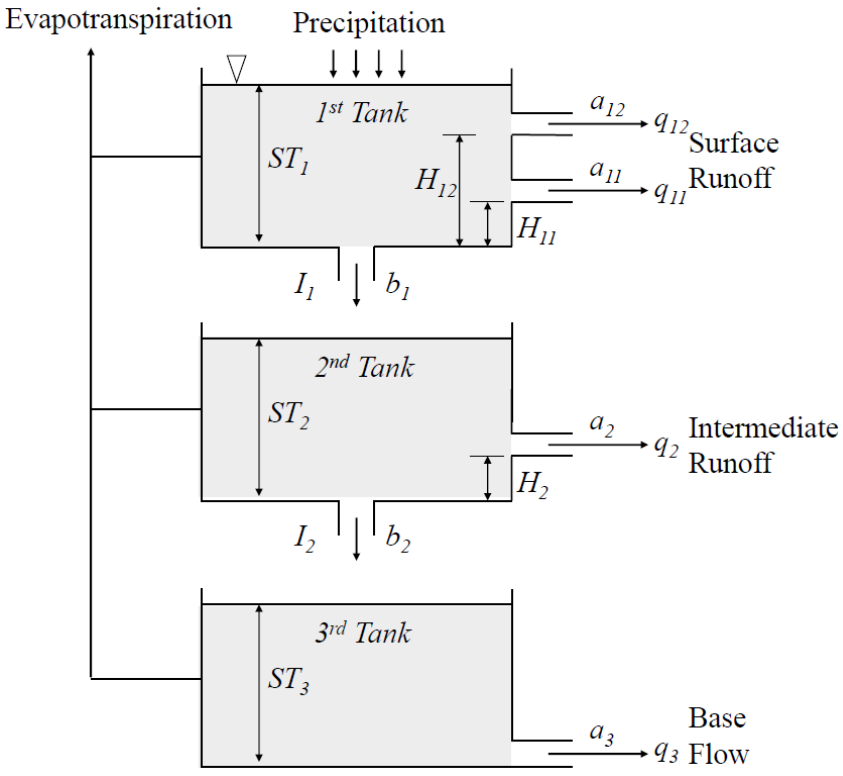


Figure 6-1 Schematic diagram showing the structure of the 3-Tank model.

The tank depth is directly recharged by rainfall in the first tank and through infiltration or percolation in the remaining tanks and depleted through either evapotranspiration or runoff so that

the amount of water in each tank represents catchment storage. In terms of the 3-Tank model, a continuous daily rainfall is used in the first tank as input and evapotranspiration is assumed to occur in all three tanks. For instance, if there is insufficient water in the first tank, the storage of the second tank is reduced to fulfil the demand for evapotranspiration in the first tank. Similarly, the lack of water in both the top two tanks affects the storage reduction of the third tank located at the bottom. The side outlets in each tank are responsible for simulating runoff in different layers, representing surface runoff, intermediate runoff, and base flow, respectively. Here, total runoff is then calculated by accumulating runoff from the side outlets as follows (Song et al., 2017):

$$Q = q_{11} + q_{12} + q_2 + q_3 = \sum(ST_i - H_{ij})a_{ij} \quad (6\text{-}1)$$

where Q is total runoff and q_{ij} refers to the runoff of j^{th} side outlet at the i^{th} tank (mm) with the associated runoff coefficient a_{ij} while ST_i represents the storage of the i^{th} tank (mm). More details on the model equation can be found in Appendix A.

To accurately estimate runoff through the tank model, reliable daily rainfall and potential evapotranspiration datasets over a given catchment are required as input data. Here, daily rainfall sequences from six stations are obtained, and the areal mean rainfall over the catchment is calculated by the Thiessen polygon method. Moreover, the standard FAO-56 Penman-Monteith method is used to estimate reference evapotranspiration (Allen et al., 1998) and the potential evapotranspiration estimates are then adjusted via the calibration process in the model, considering water balance over the catchment between input (rainfall) and output (runoff).

6.3.2 Least square support vector machine

The support vector machine (SVM) developed by Vapnik (1995) has been widely used for classification and regression tasks in many different fields, including RR prediction (Okkan and Serbes, 2012; Raghavendra and Deka, 2014). In this study, the LSSVM that is considered as an upgraded version of the SVM is employed for the purpose of RR modelling. The LSSVM differs from the conventional SVM in that the LSSVM approach uses a set of linear equations for solving optimisation problems instead of using a quadratic form in the conventional SVM (Kisi and Parmar, 2016; Yan and Chowdhury, 2013). A structure of the LSSVM is schematically presented in Figure 6-2. As abovementioned, the LSSVM particularly adopts a least squares linear system as the loss

function with two-layer networks. Here, I briefly describe the concept of the LSSVM. A more detailed description of the LSSVM scheme can be found in the literature (Suykens et al., 2002).

Let $\{x_k, y_k\}_{k=1}^{N}$ be N length of datasets with input ($x \in R^N$) and output ($y \in R$), where R^n denotes n-dimensional input/output vector space. The LSSVM model can be formulated as follows:

$$y(x) = w^T \varphi(x) + b \tag{6-2}$$

where φ (.) is the feature map embedding the input data into a higher dimensional feature space; w is a weight vector; b is considered as a bias term.

For the LSSVM regression, the optimisation problem is given as follows:

Minimize: $\frac{1}{2} w^T w + \frac{1}{2} \sum_{k=1}^{N} e_k^2$,

Subject to: $y(x) = w^T \varphi(\mathrm{x_k}) + b + e_k,\ K = 1, \ldots, N$ (6-3)

Then, Eq. (6-3) is finally rewritten by introducing Lagrange multipliers as follows:

$$y(x) = \sum_{k=1}^{N} a_k K(x, x_k) + b \tag{6-4}$$

where a_k is the Lagrangian multiplier and $K(x_k, x_l)$ is the kernel function.

$$K(x_k, x_l) = \exp\left\{-\frac{(x_k - x_l)^T (x_k - x_l)}{2\,\sigma^2}\right\},\ \ k, l = 1, \ldots, N \tag{6-5}$$

where σ is the width of the radial basis function.

It should be noted that the performance of the LSSVM is highly dependent on the kernel function. Among many kernel functions, such as linear kernel function, polynomial, sigmoid and radial basis function (RBF), the RBF is used in the practical aspects. More specifically the RBF is more flexible than the others with fewer parameters to be estimated (Hosseini and Mahjouri, 2016). The LSSVM with Coupled Simulated Annealing (CSA) optimisation algorithm initially determines a suitable set of parameter based on five multiple starters. These parameters are then given in the second optimisation procedure to further optimise the parameters in a fine-tuning step: here, a simplex optimisation approach (Nelder and Mead, 1965) is employed for tuning parameters through a cross-validation procedure by partitioning samples into training data and test dataset. For the sake

of comparison with the Tank model under the same conditions, hydrologic data used for the calibration period (2007-2013) are exploited for the training process in the LSSVM model and the remaining data for the verification period (2014-2016) are similarly considered for the testing phase. A Matlab version of the LSSVM used in this study is obtained at https://www.esat.kuleuven.be/sista/lssvmlab/.

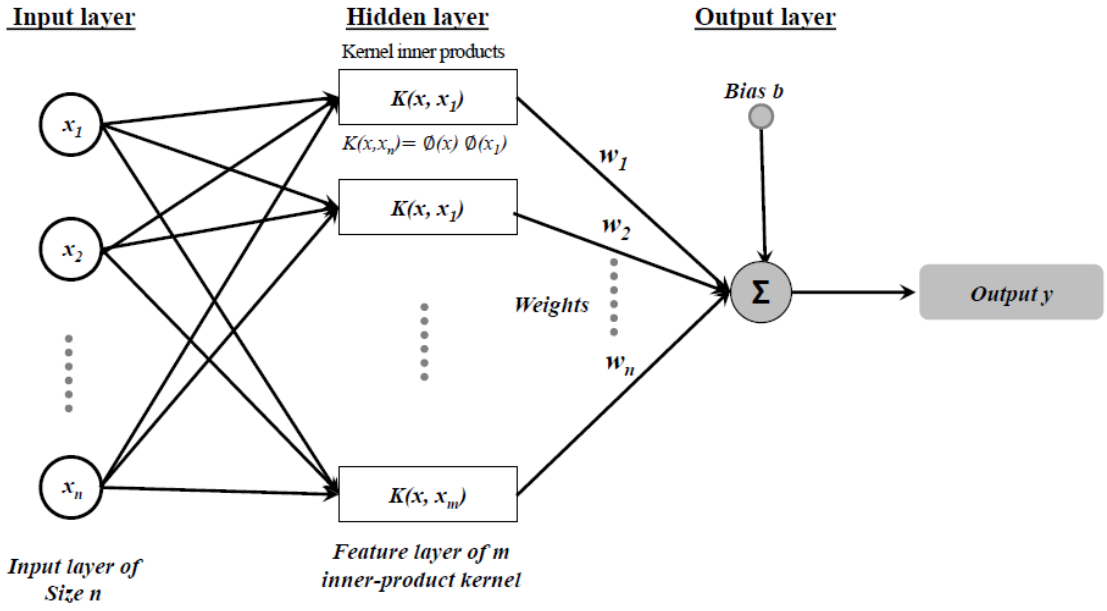


Figure 6-2 Schematic architecture of SVM based on an RBF network.

6.3.3 Tank-LSSVM hybrid model

A hybrid RR model (hereinafter named the Tank-LSSVM model) is constructed by using the outputs from the Tank model in a LSSVM based regression framework. In this study, the storage variables (i.e., ST_1, ST_2 and ST_3) in the Tank model are considered as intermediate state variables, representing the temporal variation of SM over the catchment. A schematic representation of the proposed hybrid model together with specific implementation procedures is presented in Figure 6-3. The selection of input variables (or predictors) is of great importance in a machine learning-based approach (Bray and Han, 2004). In this study, potential predictors being considered for formulating the Tank-LSSVM model are rainfall, tank storages (i.e., *ST*) and European Space Agency Climate Change Initiative (ESA CCI) SM. Moreover, the use of satellite SM products is

of particular interest for an ungauged watershed, and an experimental study is presented to explore the role within the proposed model. Here, an iterative approach to obtain the optimal combination sets of the independent variables is carried out with the use of time-lagged variables in the context of a stepwise regression. The partial autocorrelation function (PACF) is further utilised to explore the number of time-lagged variables in terms of the relative importance in the hybrid prediction model.

Prior to the construction of the LSSVM based RR model, the exploratory (independent) variables considered in this study are all normalised. The normalisation process is commonly adopted in the data-driven models with intent to reduce the problems posed by the use of relatively high values (Yu et al., 2006). Here, all the variables (i.e., independent and dependent) used in this study are equally weighted by employing a simple normalisation approach as follows:

$$Z_i = \frac{x - \bar{x}}{S} \tag{6-6}$$

where Z_i is the scaled value. $\bar{x}$ and S represent the mean and standard deviation of the original data.

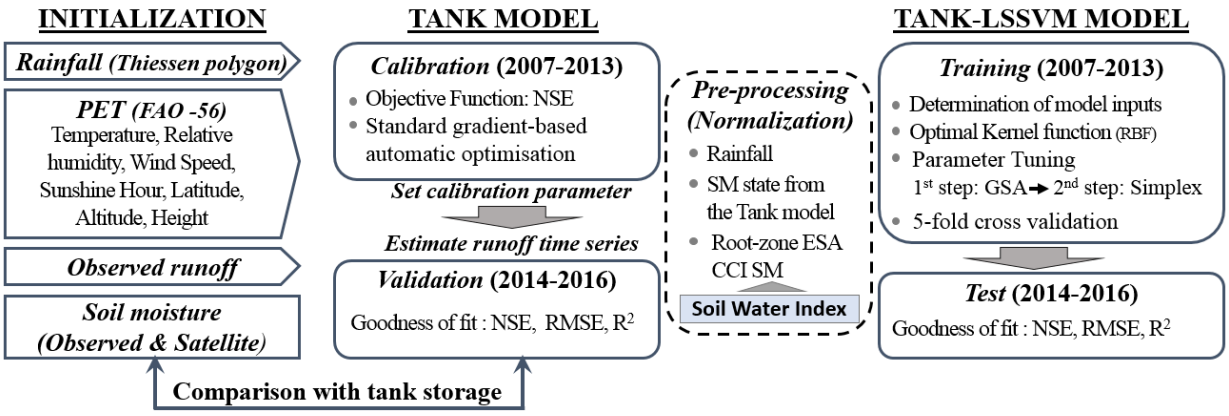


Figure 6-3 Schematic representation of the Tank-LSSVM modelling process.

6.3.4 Root-zone ESA CCI SM

As stated in Section 2.3, one fundamental issue in using remotely sensed SM products is that they are measured a few centimetres below the surface, which makes it difficult to use directly for hydrological applications (Massari et al., 2014). Moreover, RZSM contents are known to be more

representative to simulate a catchment response to rainfall (Brocca et al., 2010; Massari et al., 2015). Given the mismatch between satellite SM and the model configuration for soil layers in the Tank model, exponential filter method (also known as the soil water index (SWI); Albergel et al., 2008) was used to derive the RZSM from the original ESA CCI SM. This approach is commonly used for a pre-processing step when satellite-derived SM is integrated into hydrological models (Loizu et al., 2018). Similar to Chapter 4, the optimal characteristic time length T is obtained by maximising the correlation coefficient (r) between SWI and simulated SM (i.e., tank storage) in the Tank model. The original surface ESA CCI SM series is smoothened through this approach, which can be considered as a proxy for SM in a deeper layer.

6.3.5 Performance scores

The efficiency of the proposed RR models is evaluated using three goodness-of-fit measures (i.e., the Nash-Sutcliffe efficiency, coefficient of determination and root mean square error), which are commonly adopted in the field of hydrological and hydroclimatic models (Legates and McCabe, 1999). Among these criteria, RMSE is the widely used absolute criterion, while NSE is the normalised relative criteria that assesses the relative magnitude of the residual variance (i.e., noise) in comparison with the variance of the target output sequence (Nash and Sutcliffe, 1970). NSE recommended by ASCE (1993) is a convenient and popular indicator for evaluation of hydrological models and regarded as a classic skill score. The NSE emphasizes model performance for high flow due to the use of the observed mean as base line as well as its squared form (Song et al., 2017). For NSE < 0 the model performs no better than using the observed mean value, while acceptable NSEs are in the interval (0 - 1). A more detailed description of the three statistical measures is presented in Table 6-1. Here, RMSE Q70 is particularly adopted to quantify the model performance in low flows within the range of 70–100 % time exceedance following Pfannerstill et al. (2014). Additionally, descriptive performance criteria with NSE proposed by Kalin et al. (2010) are considered to determine the degree of accuracy in simulating daily runoff as follows: *Very good: NSE ≥ 0.7; Good: 0.5≤ NSE < 0.7; Satisfactory: 0.3≤ NSE < 0.5; Unsatisfactory: NSE< 0.3*.

Table 6-1 Performance metrics employed in this study. O and $\overline{O}$ indicates observed runoff and the mean of the observed runoff, respectively. E is the simulated runoff and $\overline{E}$ is the mean of the simulated runoff. Here, n is the number of observations.

Performance Metrics	Equations	Range	Optimal value
Nash–Sutcliffe efficiency (NSE)	$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - E_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2}$	-∞ ~ 1	1
Coefficient of determination (R^2)	$R^2 = 1 - \frac{\left[\sum_{i=1}^{n}(O_i - \overline{O}) \cdot (E_i - \overline{E})\right]^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2 \cdot \sum_{i=1}^{n}(E_i - \overline{E})^2}$	0 ~ 1	1
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - E_i)^2}{n}}$	0 ~ ∞	0

6.4 Results and discussion

In this section, I first explore the performance of the Tank model. A machine learning-based RR model (LSSVM) is then constructed using the lagged rainfall (P_{t-n}) and ESA CCI_{SWI} (θ_{t-n}). Finally, a hybrid model (Tank-LSSVM) is subsequently introduced to incorporate a set of intermediate state variables obtained from the Tank model into a LSSVM based RR modelling framework.

6.4.1 Tank model

The Tank model is a lumped conceptual RR model that requires a small amount of data, usually the rainfall, and relatively few model parameters. There are ten parameters to be estimated through a model calibration process in that the model parameters are systematically adjusted so as to approximate the observed streamflow. On the other hand, the validation process is carried out based on the calibrated parameters. The primary purpose of the validation process is to ensure that the model can be further generalized to totally unseen data and can successfully reproduce the observed streamflow on a daily basis over different periods of time. In this context, both calibration and validation are performed under different climate regimes to better evaluate model performance. Here, daily rainfall and streamflow data from 2007 to 2013, which can be regarded as the desired

climate conditions (i.e., average, wet and dry years), are selected and the remaining data (from 2014 to 2016) are used for model validation (Figure 6-4).

In this study, the optimal parameter set for the calibration period is derived based on a standard gradient-based automatic optimisation scheme (Bober, 2013). A wide range of model parameters and their fitted values via the optimisation scheme in the calibration process are summarised in Table 6-2. In Figure 6-5, the daily simulated and observed runoff Q together with rainfall, during the investigation period (2007-2016), are given in a plot of log Q, suggesting that the Tank model performance for both calibration and validation period can be regarded as "very good" in terms of the NSE. More generally, it can be concluded that runoff process can be effectively reproduced by the Tank model with an acceptable level of accuracy in terms of R^2 and RMSE (R^2 = 0.92 (0.83) and RMSE = 20.18 m^3/s (14.72 m^3/s) for calibration (validation) phase, respectively). However, the Tank model does not fully capture the complex behaviour of the RR processes in the validation phase, compared to the calibration period (i.e., 0.92 for the calibration period and 0.81 for validation phase, respectively). A notable deviation between observed and simulated runoff is especially observed during the dry period (i.e., recession curve). The results of the present study suggest that the sole use of the Tank model may have limited performance in effectively describing the dynamics of low flow. In this perspective, this study seeks a more rigorous approach to the issue of low flow simulation in a hybrid modelling framework.

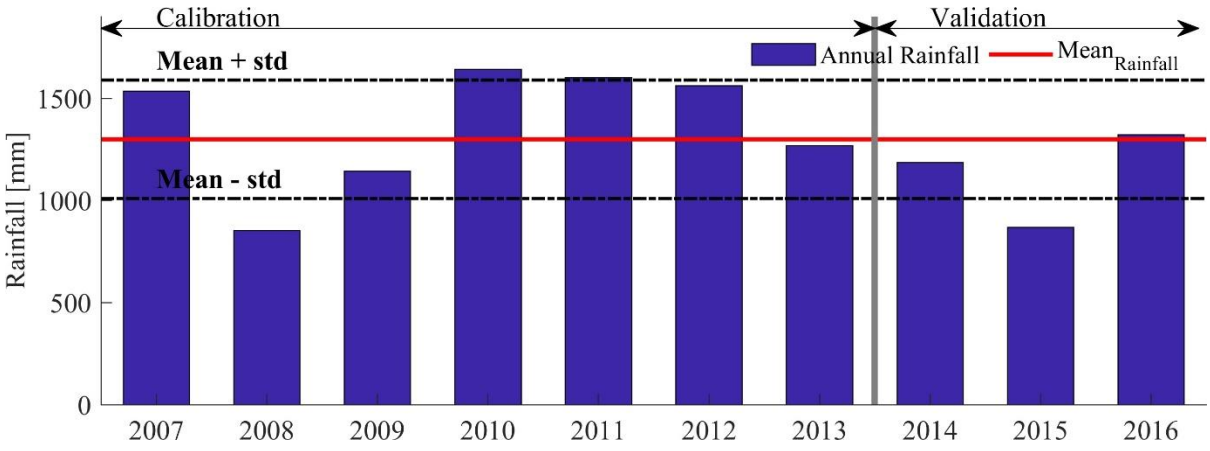


Figure 6-4 Temporal distribution of annual rainfall over the Yongdam catchment during calibration and validation phases considered in this study. The red solid line and black dotted lines represent the mean and 1 standard deviation, respectively.

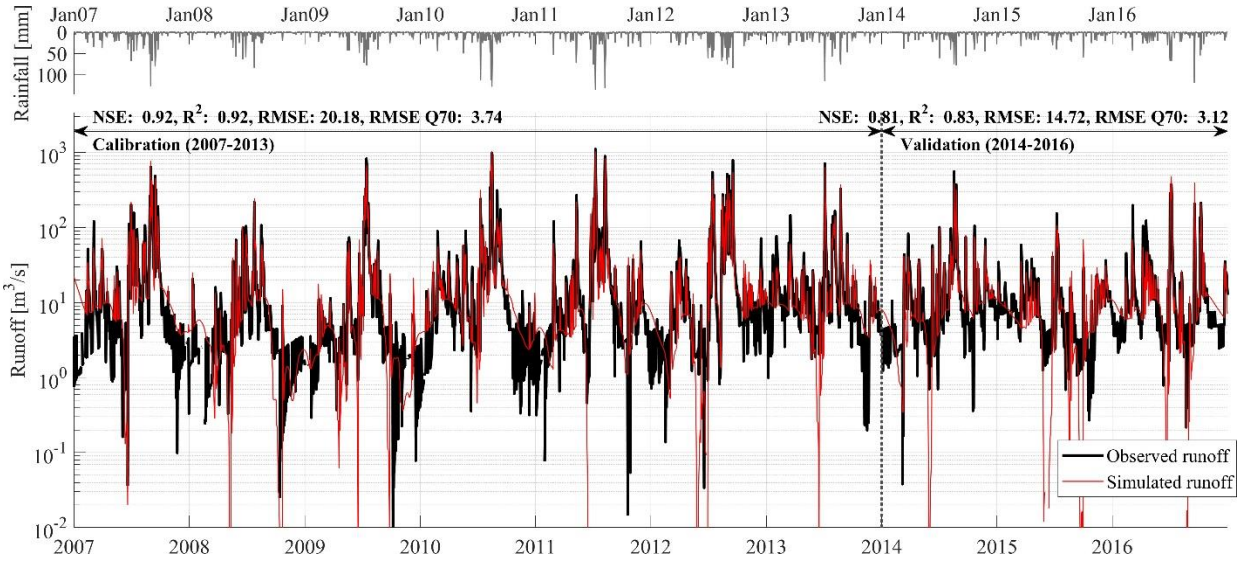


Figure 6-5 Time series plot showing observed runoff and runoff simulated by the Tank model. The upper panel depicts the average catchment rainfall over the course of the investigation period.

Table 6-2 Parameter range and the optimum values estimated through the calibration process.

Parameter		a11	a12	a2	a3	b1	b2	b3	h11	h12	h20
Range	Min.	0.08	0.08	0.03	0.00	0.10	0.01	0.00	5.00	20.00	0.00
	Max.	0.50	1.00	1.00	0.03	0.50	0.35	0.11	60.00	150.00	100.00
Obtained value		0.13	0.33	0.71	0.02	0.14	0.07	0.01	10.72	62.94	35.14

This study aims to explore a hybrid model that incorporates the obtained intermediate variables (i.e. tank storage) from the Tank model into the LSSVM-based RR modelling framework. To begin with, I explore the correlations between tank storages and in-situ SM to determine the effectiveness of tank storages as a proxy measure of the SM, representing the temporal variability of SM over the catchment. Here, six different cases are explored to find how the individual tanks (i.e. Case 1-3) and their combinations (i.e. Case 4-6) are correlated with in-situ SM, as seen in Figure 6-6.

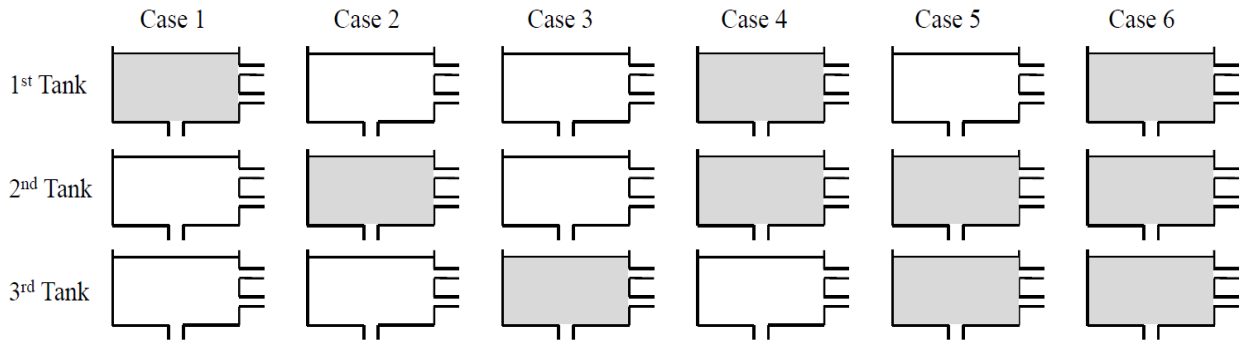


Figure 6-6 Schematic representation of six different tank combinations. Grey-coloured tanks refer to the tanks used in comparison with in-situ SM observations.

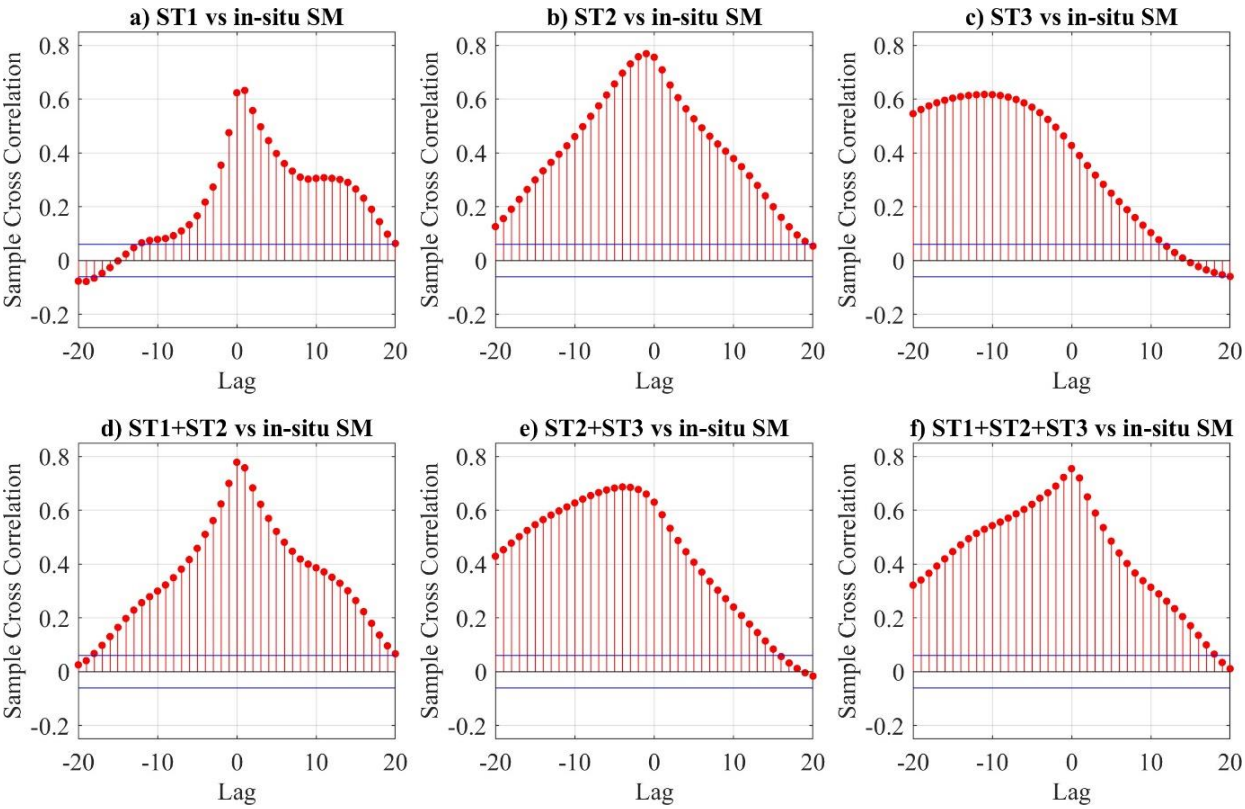


Figure 6-7 Lagged cross-correlation between tank storages and in-situ observations during the period 2014-2016. The solid blue line represents the 95%-confidence interval. The value outside of the confidence interval is statistically significant.

As stated in Section 3.2.4, it should be noted that in-situ SM observations in the study area are found to be limited to relatively small periods, ranging from 2014 to 2016. Here, I analyse the lagged windowed cross-correlation to quantify temporal coherence between tank storages and in-situ SM that is spatially averaged over six measurement points, as shown in Figure 6-7. It is clearly seen that there is a statistically significant correlation at the 95% confidence level. For the zero-lag correlation between individual tank storage and in-situ SM, the most robust temporal coherence is observed in the 2nd tank with the correlation coefficient (r) of 0.77, while the lowest r value (0.43) is shown in Case 3, suggesting that the obtained intermediate variables from each tank to a certain extent describe the temporal dynamics of SM. Interestingly, the combinations of tank storages (i.e., Case 4-6) demonstrate slightly higher or lower r values compared to the case 2 (the 2nd tank storage). In terms of time lag, Case 2 shows the highest r value of 0.78 with lag-1, while for the 3rd tank the strongest correlation is observed with a lag-11, which may be due to a slow varying behaviour of SM. Overall, the results obtained from the cross-correlation analysis confirm that the temporal dynamics of the tank storages are closely related to the variation of SM, suggesting that the intermediate variables from the Tank model can be used as a proxy for the SM. In the following section, I will further explore the role of the intermediate variables, particularly for low flow simulation within a machine learning based RR model.

6.4.2 LSSVM and Tank-LSSVM model

6.4.2.1 Determination of model inputs

The performance of a machine learning based hydrological model is considerably dependent on the choice of lagged input vectors for model training (Bray and Han, 2004; Yu et al., 2006). In this study, apart from rainfall and satellite SM, the tank storages derived from the Tank model are considered as state variables for a proxy of SM content. Additionally, time-lagged relations between hydrologic variables are taken into account, which enables one to consider a sequential hydrological process in the proposed framework. The input variables and time lags appear to be more important, but their selection procedures are rather ad hoc process in the machine learning based regression approach (Bray and Han, 2004). In other words, there is no universal and generalized model to directly select the optimal time-lagged input vector; thus an empirical approach to the selection procedure is favourable.

In this sense, I first use the partial autocorrelation function (PACF) to gain insights with respect to the smallest time lags for a parsimonious model, which an autoregressive memory holds. The plots of the PACF analysis for each input variable are presented in Figure 6-8. Based on the results obtained from the PACF analysis, the extent of lags for each variable is determined and the lagged input vectors are further considered as candidate predictors in the proposed hybrid model. For instance, a statistically significant PACF in ST1 is observed by lag-2 and the PACF falls outside the 95% confidence interval at lag-3. In this sense, the lagged vectors ranging from lag-0 to lag-2 for ST1 are primarily considered, and similarly for other variables.

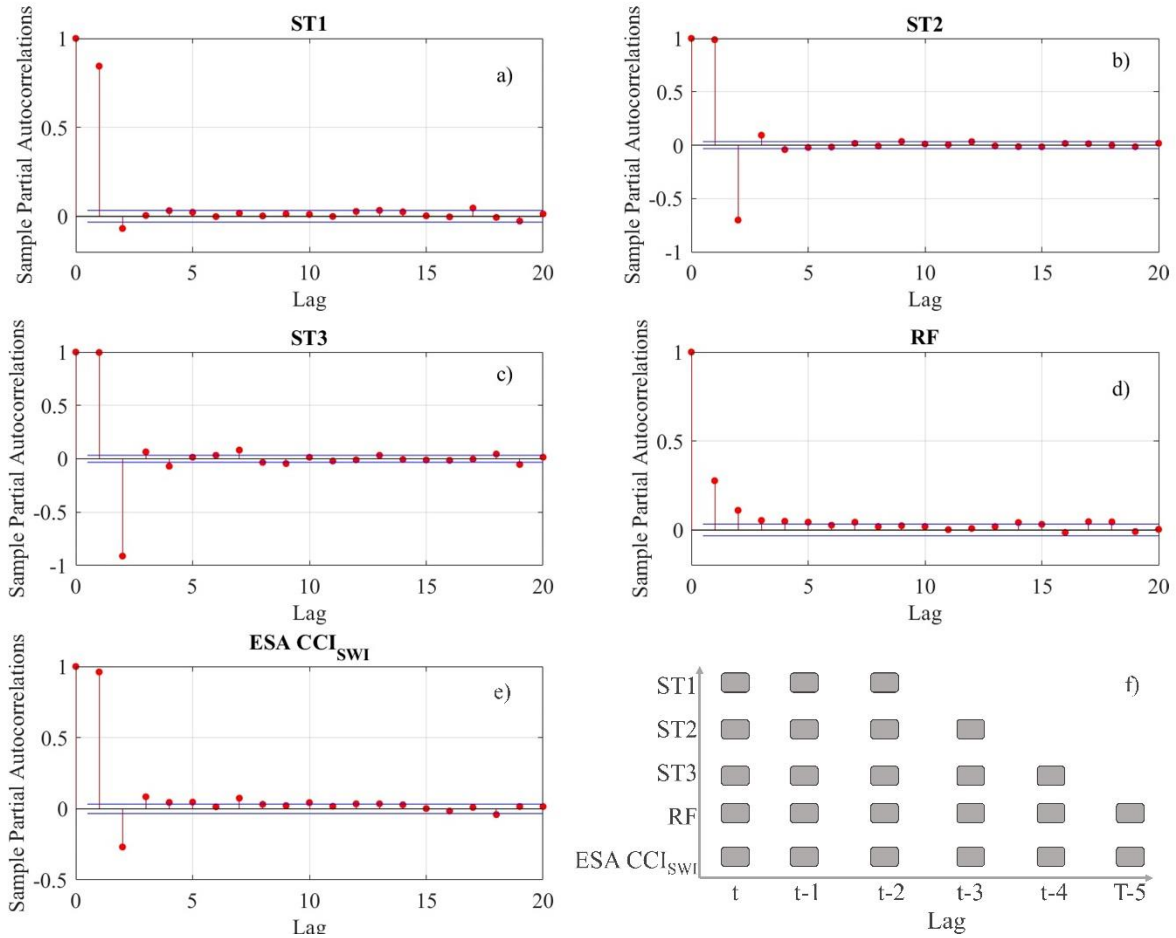


Figure 6-8 The partial autocorrelation function of input variables (a-e) and their time lags considered in this study (f). STn, RF and ESA CCI_{SWI} represent the n-th tank storage in the Tank model, rainfall and Root-zone ESACCI SM data, respectively.

6.4.2.2 LSSVM model

6.4.2.3

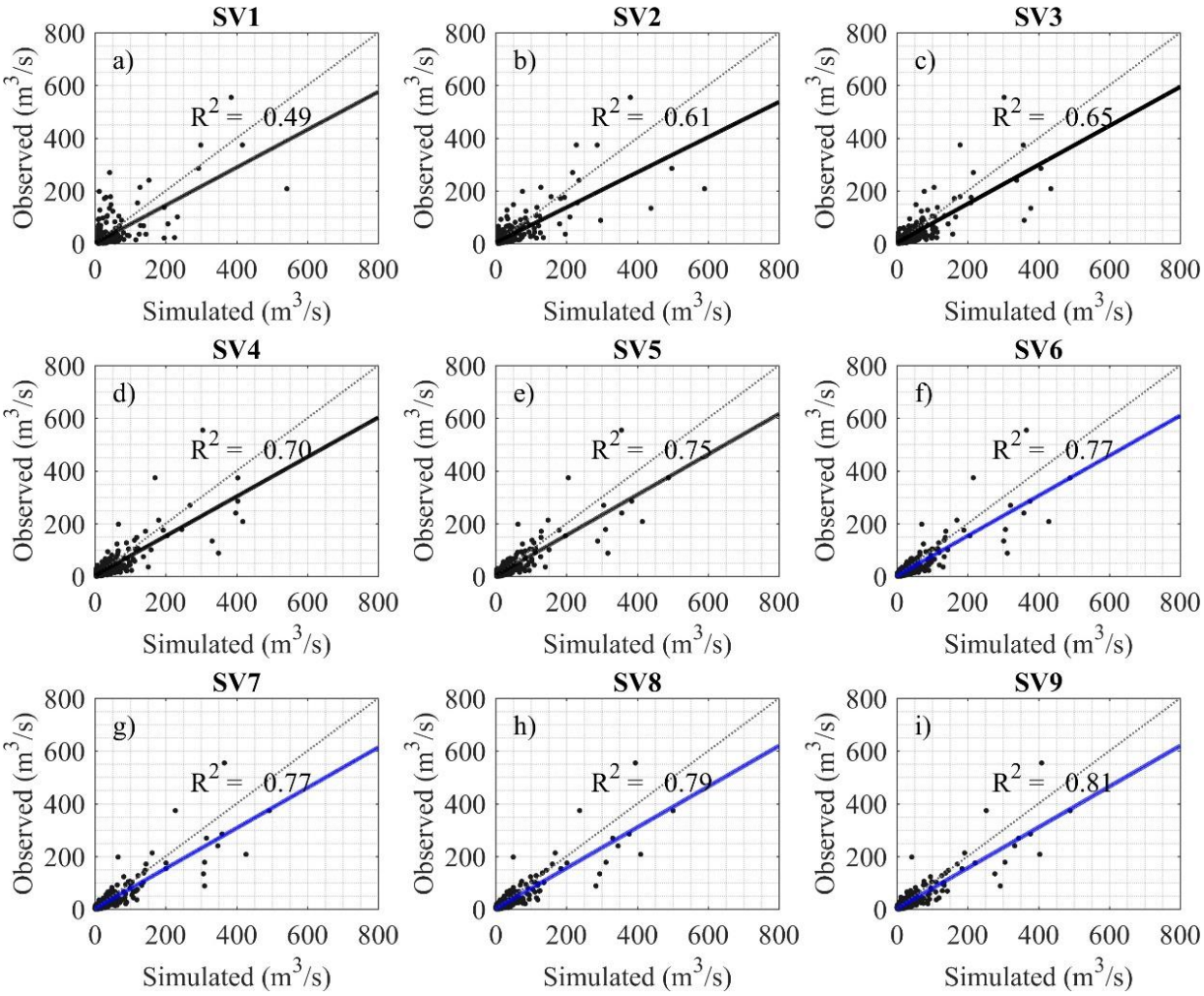


Figure 6-9 SV1-SV9 model scatter plots with corresponding linear regression lines and R^2 over the testing period.

The LSSVM-based RR model is constructed using several lagged values of the independent variables, rainfall (P_{t-n}) and ESA CCI_{SWI} (θ_{t-n}), as inputs (without a set of intermediate variables from the Tank model). Here, the input vectors are partitioned into two subsets (i.e., training and testing phase) with the same period as considered in the Tank model, which can facilitate a comparison with the results from the Tank model under the same condition. The dependent and independent variables are all normalised prior to applying the LSSVM model, using Eq. (6-6).

In this study, I first attempt to build the LSSVM-based runoff model using a single predictor of rainfall data and the ESA CCI_{SWI} data are then added as an additional input with an intention to explore the contribution of satellite-based SM in the LSSVM-based RR model. Among many hydrological variables, rainfall and satellite-based SM are particularly selected as candidate inputs due to the fact that the two variables are not only significantly related to runoff process but also closely related to each other, indirectly considering interdependence in the hydrological cycle. In Table 6-3, the results of runoff simulation with different combinations of input variables are presented. The relative impact of different combinations of lagged input variables on the model performance is investigated in a stepwise manner by repeatedly adding more lagged values until no improvement can be found.

I mainly focus on the improvement of performance during the testing phase. In terms of the NSE, a significant increase in performance efficiency is clearly seen (0.40 for the SV1 and 0.68 for the SV5) until lag-4 and no improvement in performance with the more lagged values. Similar results are obtained for other performance measures (i.e., an increase in R^2 from 0.49 to 0.75 and a decrease in RMSE from 26.02 to 18.91. Rainfall data from lag-0 to lag-4 are therefore used for the subsequent analysis. As summarised in Table 6-3, SV6-SV9 models additionally introduce the lagged ESA CCI_{SWI} data as inputs in the LSSVM modelling framework, showing better and comparable performance than that of the former SV1-SV5 models, except for the RMSE Q70 during the testing period. This indicates that the SM states inferred from the ESA CCI_{SWI} appear to provide an effective means of describing the temporal dynamics of SM in the LSSVM-based RR model. In terms of the NSE, a slight increase in the values, ranging from 0.69 to 0.73, is identified for the remaining SV6-SV9 models. Similar improvements in other performance measures (i.e. R^2, RMSE and RMSE Q70) are observed. Based on the performance criteria, the SV8 and SV9 models can be regarded as "very good" performances (NSE > 0.7). Unlike other measures, no improvement in the RMSE Q70 is seen in the cases including ESA CCI_{SWI} data, suggesting that there is a limited role for the direct use of remote sensing SM products in the context of low flow simulation. The performance of simulated runoff from the LSSVM models is compared to the observed runoff during the testing period through scatter plots with the corresponding R^2. In terms of the R^2, enhanced results in the SV6-SV9 models are clearly demonstrated (Figure 6-9).

Table 6-3 Model performance measures with different combinations of lagged input variables during the training and testing periods. Here, p and θ represents rainfall and ESA CCI_{SWI}, respectively.

Model	Input combinations	Training (2007-2013)				Testing (2014-2016)			
		NSE	R^2	RMSE	RMSE Q70	NSE	R^2	RMSE	RMSE Q70
SV1	$P_{(t)}$	0.60	0.60	44.72	10.40	0.40	0.49	26.02	9.86
SV2	$P_{(t)}$, $\theta_{(t-1)}$	0.84	0.84	28.21	6.73	0.45	0.61	24.81	5.42
SV3	$P_{(t)}$, …, $P_{(t-2)}$	0.88	0.88	24.88	5.11	0.57	0.65	22.04	4.49
SV4	$P_{(t)}$, …, $P_{(t-3)}$	0.89	0.89	23.01	4.44	0.62	0.70	20.77	3.99
SV5	$P_{(t)}$, …, $P_{(t-4)}$	0.91	0.91	21.33	4.00	0.68	0.75	18.91	3.86
SV6	$P_{(t)}$, …, $P_{(t-4)}$, $\theta_{(t)}$	0.91	0.91	21.24	3.18	0.69	0.77	18.80	4.79
SV7	$P_{(t)}$, …, $P_{(t-4)}$, $\theta_{(t)}$, $\theta_{(t-1)}$	0.91	0.91	20.74	3.29	0.69	0.77	18.62	5.02
SV8	$P_{(t)}$, …, $P_{(t-4)}$, $\theta_{(t)}$, …, $\theta_{(t-2)}$	0.92	0.92	19.53	3.12	0.72	0.79	17.90	4.93
SV9	$P_{(t)}$, …, $P_{(t-4)}$, $\theta_{(t)}$, …, $\theta_{(t-3)}$	0.92	0.92	19.33	2.98	0.73	0.81	17.51	4.75

6.4.2.4 Tank-LSSVM model

Here, a hybrid model is introduced by bridging the intermediate state variables from the Tank model and a machine learning framework (the LSSVM). In other words, several lagged values of tank storages (ST_{t-n}) derived from the Tank model are subsequently utilised in the RR simulation model, together with rainfall and ESA CCI_{SWI} used for constructing the LSSVM model in the previous section. The results of RR simulations with different combinations of input variables are summarised in Table 6-4. Herein, the results obtained from the Tank model are also presented. The HY1-HY4 models are carried out without consideration of time lags, whereas the HY5-HY9 models consider time-lagged input vectors. Contrary to the LSSVM model, the use of the time-lagged input variables shows little or no improvement in the performance when the intermediate SM state variables from the Tank model are used in the Tank-LSSVM model (i.e., the HY5-HY9 models). This motivates the need for extensive comparisons over the HY1-HY4 model. The HY1 model considers only the first tank storage and rainfall as input variables, while the 2nd and 3rd tank are further included in the HY2 and HY3 model, respectively. As for the HY4 model, ESA CCI_{SWI}

data sets are combined into the HY3 model for the purpose of assessing the contribution of satellite SM to the hybrid RR modelling framework.

Table 6-4 Tank-LSSVM model performance measures using different input combinations. Here, p and θ represents rainfall and ESA CCI_{SWI}, while STn refers to the n-th tank storage.

Model	Input combinations	Training (2007-2013)				Testing (2014-2016)			
		NSE	R^2	RMSE	RMSE Q70	NSE	R^2	RMSE	RMSE Q70
Tank		0.92	0.96	20.18	3.74	0.81	0.91	14.72	3.12
HY1	$P_{(t)}$, $ST1_{(t)}$	0.92	0.96	19.74	4.19	0.75	0.80	16.63	3.11
HY2	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$	0.93	0.96	18.49	2.99	0.76	0.80	16.52	2.67
HY3	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$, $ST3_{(t)}$	0.95	0.97	16.24	2.50	0.85	0.86	12.91	2.13
HY4	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$, $ST3_{(t)}$, $\theta_{(t)}$	0.94	0.97	17.43	2.34	0.85	0.85	12.96	2.23
HY5	$P_{(t)}$, $P_{(t-1)}$, $ST1_{(t)}$	0.92	0.96	19.87	4.40	0.71	0.77	17.93	3.11
HY6	$P_{(t)}$, $ST1_{(t)}$, $ST1_{(t-1)}$, $ST2_{(t)}$, $ST3_{(t)}$	0.93	0.96	18.75	2.94	0.84	0.85	13.38	2.19
HY7	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$, $ST2_{(t-1)}$, $ST3_{(t)}$	0.93	0.96	18.87	2.99	0.85	0.85	13.18	2.34
HY8	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$, $ST3_{(t)}$, $ST3_{(t-1)}$	0.93	0.96	18.97	2.93	0.85	0.85	13.13	2.39
HY9	$P_{(t)}$, $ST1_{(t)}$, $ST2_{(t)}$, $ST3_{(t)}$, $\theta_{(t)}$, $\theta_{(t-1)}$	0.93	0.96	18.98	2.81	0.85	0.85	13.14	2.49

In terms of NSE, the Tank-LSSVM models are all classified as "very good" based on the performance criteria during the training and testing periods, with different levels of accuracy. Improved modelling of the daily runoff simulation is clearly seen in the Tank-LSSVM models that include a set of state variables from the Tank model, in comparison to the well-calibrated individual LSSVM and the Tank model. The overall performance of the HY1 model is comparable to that of the HY2 model but a significant improvement in low flow simulation in terms of the RMSE Q70 is evident during both the training and testing period in the HY2 model (i.e., 2.67 m^3/s), compared to the Tank model (i.e., 3.12 m^3/s). It can be concluded that the intermediate SM states inferred from the 2nd tank storage ST2 play an important role in the simulation of low flow in the HY2 model. Moreover, the inclusion of the 3rd tank storage in the HY3 model significantly improves the low flow simulation, in terms of the RMSE Q70 (i.e., 2.13 m^3/s), with comparable results over other performance metrics from the two models (i.e., the HY1 and HY2). In a similar context of the HY2 model, it seems that the 3rd tank storage plays a critical role in describing the

base flow as a proxy variable, which accordingly leads to a significant improvement in the simulation of low flow.

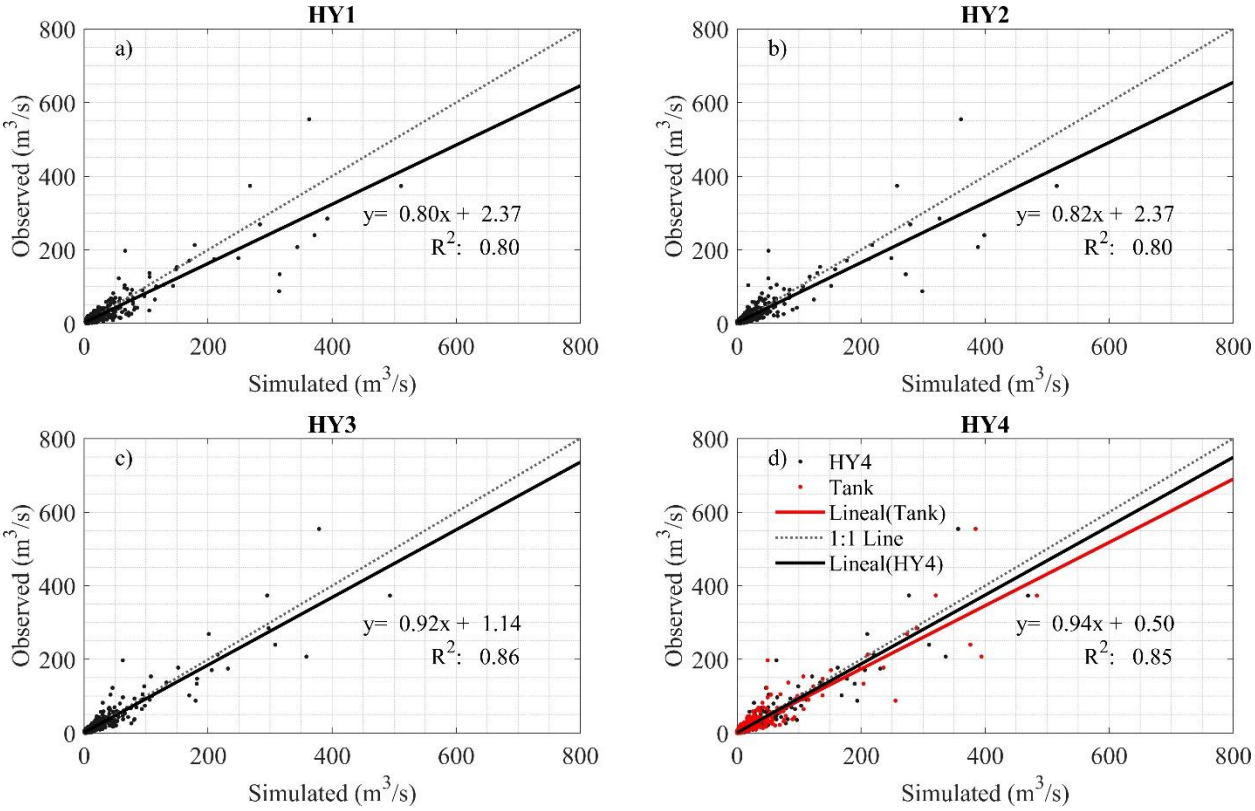


Figure 6-10 Scatterplots of the observed and simulated runoff during the testing period. The dotted lines represent perfect linear relationships while the solid black lines indicate a linear fit between the simulated and observed runoff.

The performances of the Tank-LSSVM models are also confirmed by a graphical representation, as displayed in Figure 6-10. As for the linear regression, the simulation of the HY4 model provides approximately a 6 % underestimation of the observed runoff (i.e., $y = 0.94\ x + 0.5$), showing the best performance among other hybrid models including the Tank model. However, the contribution of satellite SM products in the Tank-LSSVM seems to be insignificant. This is partially due to the fact that the temporal dynamics of SM are entirely described by the tank storages, although the HY4 model provides a better performance in terms of the RMSE Q70 during the training period.

It should be noted that accurate RR modelling during dry seasons is of great importance for water resources planning and management studies, especially for problems associated with low flows such as water quality and drought as well as water budget analysis. To further assess the accuracy of low-flow simulations, the flow duration curves derived from simulated runoff series are graphically compared against those of the observed during the testing phase (2014-2016). Apart from the HY1 model that only uses the first tank storage, the performances for low flow simulation in the remaining Tank-LSSVM models are remarkably better than those of the Tank model, as illustrated in Figure 6-11.

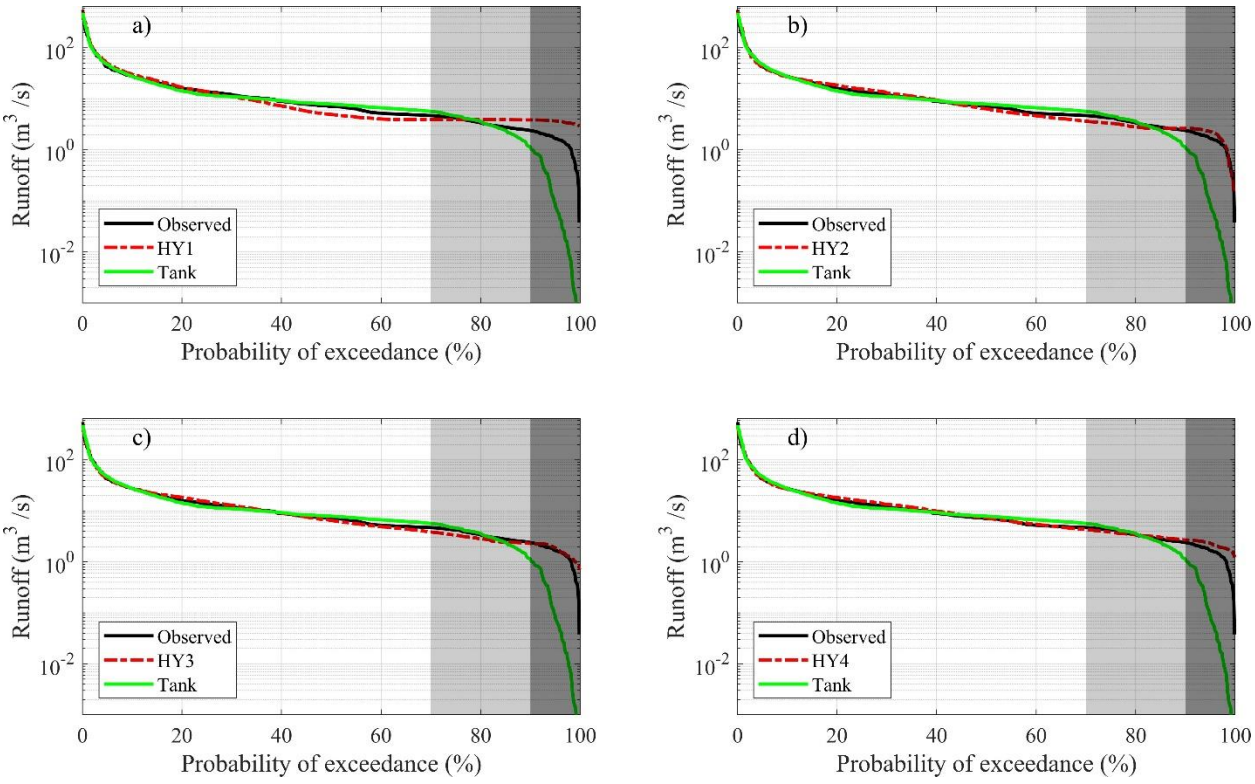


Figure 6-11 Flow duration curve comparison between four different Tank-LSSVM models, the Tank model, and direct observation data during the testing period (2014-16).

Figure 6-12 shows the simulated runoff time series obtained from two models (the Tank model and Tank-LSSVM model) with the observed runoff, indicating that the rapid fluctuations of low flows are better captured by the Tank-LSSVM model. The results reveal the important role in the low-flow simulation of the Tank-LSSVM model.

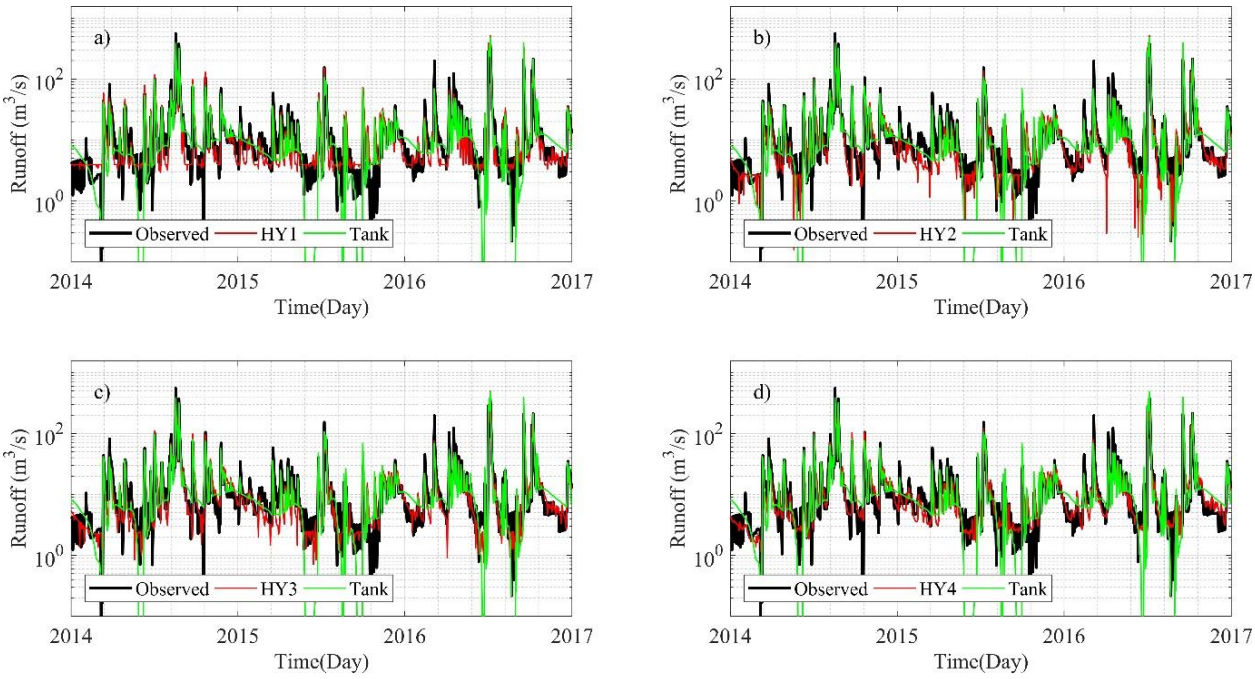


Figure 6-12 Simulated runoff comparison between the Tank model and four different Tank-LSSVM models (HY1, HY2, HY3 and HY4). Direct observation runoff data from 2014-2016 is also included.

6.5 Conclusions

In this chapter, I have explored a new RR model that combines the intermediate state variables obtained from a Tank model with a LSSVM based nonlinear regression model. The main assumption of this study was that the combination of different models could be more favourable and efficient in picking up different aspects of runoff modelling than the individual use. The performances of the hybrid RR model were compared against those of two individual RR models (i.e., the Tank model and LSSVM model). The main findings of this study can be summarised as follows:

(1) The performance of the Tank model with the calibrated parameters was first examined. The results confirmed that the Tank model is capable of accurately describing the rainfall-runoff relationships and can be categorised "very good (NSE > 0.7)" according to a level of accuracy in terms of the NSE. However, the Tank model showed relatively large deviations for a low-flow

simulation, suggesting that the sole use of the Tank model is insufficient to simulate particular features of the RR processes.

(2) The LSSVM-based RR model was introduced using two lagged values of predictors (i.e., rainfall and ESA CCI_{SWI}) as inputs. The results obtained from the proposed approach demonstrated that the satellite SM appears to be more or less effective in describing the temporal dynamics of SM. The performances of the LSSVM model are classified as "good ($0.5 \leq$ NSE <0.7)" or "very good", which depends on different combinations of the time-lagged input variables. Although the overall performances are generally lower than those of the Tank model, the results from the LSSVM model support the potential use of the satellite-based SM products in hydrological applications. To be specific, in the practical aspects, this approach appears to be an alternative to the conventional RR model in regions where the satellite-based SM products are the only measurements available.

(3) Finally, the potential use of remotely sensed SM products and their combined use with the intermediate state variables estimated through hydrological modelling were explored in the context of the hybrid simulation. It was found that the Tank-LSSVM models including the intermediate state variables lead to a significant improvement in the simulation of low flows during both the training and testing periods, in comparison to the well-calibrated individual LSSVM and Tank model, with comparable results over other performance metrics. For the use of satellite SM products in the model, the contributions were found to be insignificant in the simulation of low flows due to the fact that the temporal dynamics of SM are largely described by the tank storages. The results confirm that SM state variables derived from the well-calibrated continuous RR model can better represent the temporal dynamics of SM than those obtained from satellite SM data.

CHAPTER 7 [1]Exploration of drought features by combining soil moisture and rainfall deficits

7.1 Motivation

As mentioned in Section 2.5, drought is a periodic phenomenon and occurs virtually everywhere in the world, but its characteristic (e.g., duration, intensity and frequency) varies significantly depending on climate regimes (Mirabbasi et al., 2013). In South Korea, extreme weather events including droughts have been becoming much more frequent since the late 1990s (Yoo et al., 2012). Recently, the Korean peninsula experienced a severe drought for two years in a row (2013–2015), and the probability of drought occurrence is expected to be higher due to the delayed emergence of monsoon as well as its weakened duration (Li et al., 2017). The amount of rainfall during this dry periods was 35–50% of the annual average (1973 to 2015) and local governments implemented to restrict water usage in many cities across South Korea (Kwon et al., 2016). In this context, understanding the characteristics of drought and creating a comprehensive drought warning system that conveys drought information to public stakeholders or decision makers are urgently required to prevent and mitigate the negative effects of drought.

Considering that drought is a multidimensional phenomenon, using a drought index based on a single indicators (e.g., precipitation and SM) is likely to be insufficient for defining droughts (Hao and Singh, 2015). In this sense, combining multiple hydro-meteorological variables is beneficial for a better understanding of drought characterisation, and particularly useful for communication purposes between different types of drought. However, those aspects have rarely been investigated in South Korea due to a lack of long-term measured observations. Alternatively, here the ERA-Interim SM data are exploited to identify SM deficiency and subsequently combined with the deficit of rainfall in an attempt to provide the integrated drought information.

---

[1]Moonhyuk Kwon, Hyun-Han Kwon and Dawei Han. "Spatio-temporal Drought Patterns of Multiple Drought Indices Based on Precipitation and Soil Moisture: A case study in South Korea." *International Journal of Climatology*. DOI: 10.1002/joc.6094.

To be specific, I adopt the Copula-based multivariate standardized drought index by combining meteorological and agricultural droughts, and the proposed drought index is then grouped by the hierarchical agglomerative clustering approach for classifying regional patterns. Drought is somewhat different from other water-related hazards in terms of its spatio-temporal characteristics, resulting in structured spatial coverage with varying durations. The spatio-temporal drought patterns may differ substantially by drought intensity. In these contexts, an exploration of the spatio-temporal drought patterns over different quantiles (i.e. severity) can serve as a basis to understand the evolution and nature of droughts in space and time. However, most of the existing studies on droughts have not specifically analysed the spatio-temporal patterns at different quantiles. Thus, this study will focus on exploring the underlying structure of drought occurrence and development.

The main objectives of this chapter are fourfold: (1) To explore drought identification and relationship between the SPI and SSI indices, (2) to use a multivariate standardized drought index based on a copula function method on the basis of SPI and SSI; (3) to propose quantile regression model-based spatio-temporal drought analysis at different quantiles, and (4) to classify spatio-temporal drought patterns using the multivariate drought index and the hierarchical agglomerative clustering approach, covering the period 1986-2016 across South Korea.

7.2 Data and drought indices

7.2.1 Precipitation and SM Datasets

The historical daily precipitation data measured at 55 weather stations over South Korea, which are operated by the KMA (https://web.kma.go.kr/eng/), are collected for the period 1986-2016. Figure 3-1 shows the locations of weather stations used in this chapter and their characteristics are presented in Table 7-1. Additionally, global SM datasets from the ECMWF are used for estimating the deficit of SM. As stated in Section 3.2, the ECMWF releases global reanalysis SM datasets at four different depths (i.e., 0–7, 7–28, 28–100 and 100–289 cm). The RZSM has a significant impact on crop yield so that it is evident that crop growth and root development should be taken into consideration in designing agricultural drought indices (Narasimhan and Srinivasan, 2005).

In this respect, ERA-Interim SM data at the third layer (28-100 cm) are mainly used as the best proxy of the RZSM in this Chapter.

The reason for utilising the ERA-Interim in historical drought analysis is related to the data availability. Namely, as for drought assessment, it is recommended that one should use at least 20-30 years of historical records according to the WMO, (2012). Among the SM datasets that are taken into consideration in this thesis, the ERA-Interim solely provides more than 30 years of data, i.e., other approaches to obtaining SM information fail to meet the minimum length of SM data records for drought analysis (Figure 7-1). Note that both rainfall and SM data collected for this study are accumulated on a monthly basis for the subsequent study.

Figure 7-1 Timeline of SM data availability for each method. Here, the grey box represents data availability of in-situ observations obtained from the K-water and KMA.

7.2.2 SPI and SSI drought indices

The SPI has been widely used to effectively measure and detect the extent of a deficit of precipitation, providing locally specific early warnings of drought (Clayton, 1978). Its popularity stems from its flexibility and ease of use for detecting droughts at multiple time scales (Ganguli and Ganguly, 2016). Since the SPI was designed to provide a dimensionless index, SPI values can often be used to spatiotemporally compare an overall view of the drought at a national or global scale for a range of practical applications (Djerbouai and Souag-Gamane, 2016). To compute the SPI, daily precipitation data is first aggregated at different timescales (e.g., 3, 6, 12, 24 or 36 months).

Table 7-1 Rainfall stations used in this study along with their seasonal rainfall (winter [Nov-Jan (NDJ)], spring [Feb-Apr (FMA)], summer [May to July (MJJ)] and fall [Aug-Oct (ASO)]).

Sta. No	Sta. Name	Lat. (N)	Lon. (E)	Alt. (m)	Rainfall (mm)				
					Annual	NDJ	FMA	MJJ	ASO
90	Sokcho	38.25	128.56	18.1	1,386	142	207	653	384
100	Daegwallyeong	37.69	128.76	772.6	1,765	150	274	897	444
101	Chuncheon	37.90	127.74	76.5	1,356	70	205	850	231
105	Gangneung	37.75	128.89	26.0	1,451	154	222	653	422
108	Seoul	37.57	126.97	85.8	1,453	69	214	919	251
112	Incheon	37.48	126.62	68.2	1,237	64	194	742	238
114	Wonju	37.34	127.95	148.6	1,346	73	203	819	252
119	Suwon	37.27	126.99	34.1	1,327	73	207	803	244
127	Chungju	36.97	127.95	116.3	1,239	75	199	721	244
129	Seosan	36.78	126.49	28.9	1,273	91	219	714	249
130	Uljin	36.99	129.41	50.0	1,155	133	197	508	318
131	Cheongju	36.64	127.44	58.7	1,242	83	203	717	240
133	Daejeon	36.37	127.37	68.9	1,366	97	225	792	252
135	Chupungnyeong	36.22	127.99	243.7	1,897	136	484	931	346
138	Pohang	36.03	129.38	2.3	1,850	144	479	906	322
140	Gunsan	35.99	126.71	23.2	1,554	97	282	873	302
143	Daegu	35.89	128.62	53.5	1,233	80	258	673	222
146	Jeonju	35.82	127.15	61.4	1,298	77	236	747	237
152	Ulsan	35.56	129.32	83.2	1,309	88	230	744	247
156	Gwangju	35.17	126.89	72.4	1,091	68	199	604	220
159	Busan	35.10	129.03	69.6	1,073	69	210	589	205
162	Tongyeong	34.85	128.44	32.3	1,078	107	202	505	263
165	Mokpo	34.82	126.38	38.0	1,015	59	189	576	191
168	Yeosu	34.74	127.74	64.6	1,185	86	210	668	221
170	Wando	34.40	126.70	35.2	1,183	115	227	562	279
192	Jinju	35.21	128.12	30.2	1,243	105	211	688	240
201	Ganghwa	37.71	126.45	47.0	1,075	68	200	597	210
202	Yangpyeong	37.49	127.49	48.0	1,293	105	218	740	231
203	Icheon	37.26	127.48	78.0	1,277	109	273	631	264
211	Inje	38.06	128.17	200.2	1,394	120	241	786	247
212	Hongcheon	37.68	127.88	140.0	1,520	119	383	757	262
221	Jecheon	37.16	128.19	259.8	1,463	113	384	705	260
226	Boeun	36.49	127.73	175.0	1,163	110	231	595	228
232	Cheonan	36.78	127.12	81.5	1,445	96	351	750	248
235	Boryeong	36.33	126.56	15.5	1,536	127	384	737	288
236	Buyeo	36.27	126.92	11.3	1,524	107	340	803	274
238	Geumsan	36.11	127.48	170.4	1,337	58	210	818	251
243	Buan	35.73	126.72	12.0	1,409	68	202	893	246
244	Imsil	35.61	127.29	247.9	1,361	74	215	810	263
245	Jeongeup	35.56	126.87	69.8	1,197	60	189	729	218
247	Namwon	35.41	127.33	132.5	1,362	65	209	843	245
260	Jangheung	34.69	126.92	45.0	1,412	82	231	847	253
262	Goheung	34.62	127.28	53.1	1,308	89	219	767	233
272	Yeongju	36.87	128.52	210.8	1,238	78	191	718	251
273	Mungyeong	36.63	128.15	170.6	1,230	93	203	690	244
277	Yeongdeok	36.53	129.41	42.1	1,358	97	235	773	254
278	Uiseong	36.36	128.69	81.8	1,289	98	220	746	225
279	Gumi	36.13	128.32	48.9	1,222	113	211	661	237
281	Yeongcheon	35.98	128.95	93.8	1,352	107	226	774	244
284	Geochang	35.67	127.91	226.0	1,327	124	226	722	254
285	Hapcheon	35.57	128.17	32.0	1,348	103	229	772	243
288	Miryang	35.49	128.74	11.2	1,485	106	310	804	265
289	Sancheong	35.41	127.88	138.1	1,448	103	355	735	254
294	Geoje	34.89	128.60	45.4	1,322	75	255	748	244

In this thesis, I primarily focus on the SPI at 3- and 6-month timescales (hereinafter, SPI-3/6) to investigate the characteristics of meteorological droughts and their spatio-temporal patterns. As stated in Section 2.5, the aggregated precipitation data are typically fitted to theoretical distribution functions such as the gamma and Pearson type III distributions (Farahmand and AghaKouchak, 2015). The SPI is then computed by transforming the cumulative probability distribution into standardized normal variates with zero mean and standard deviation equal to one (Guttman, 1999; Mckee et al., 1993). However, because the optimal probability distribution of rainfall can vary substantially, a parametric approach is less flexible, leading to inconsistent results (Farahmand and AghaKouchak, 2015; Kumar et al., 2016; Vidal et al., 2009). In other words, the SPI values are inherently susceptible to the selection of a distribution function.

Therefore, I employ a non-parametric kernel density estimation approach to reduce the sampling error associated with the choice of distribution functions. While the SPI is mainly used to identify meteorological drought, agricultural drought is generally represented by the SM deficit. Accordingly, the SSI, known as an agricultural drought index monitoring the extent and degree of SM, plays a complementary role in a comprehensive review of drought conditions. Similarly, the kernel density estimation approach was used to transform SM data into the SSI. In this study, I extracted information on the durations and severities (i.e., deficit volumes) from SPI and SSI time series. Drought duration refers to the periods of the continuously negative phase, whereas drought severity is the sum of cumulative deficits over the corresponding duration (Kwon et al., 2016). Table 7-2 shows the SPI drought criteria defined by Mckee et al. (1993). Note that the same drought severity categories are subsequently applied to the SSI and MSDI.

Table 7-2 The SPI (SSI) drought severity classification and interpretation.

SPI values	Drought category
≥ 2.0	Extremely wet
1.5 to 1.99	Severely wet
1.0 to 1.49	Moderately wet
0.99 to -0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
≤ -2.0	Extremely dry

7.3 Methodology

7.3.1 Quantile regression

This study aims to assess not only the overall trends of drought characteristics but also the non-Gaussian distribution of trends in drought duration, severity and frequency at various levels of quantiles. The first-order quantile regression is applied to identify temporal trends in different drought characteristics. The τ^{th} quantile regression estimate is computed by minimizing Eq. (7-1) as follows:

$$min \sum_{i:y_i \geq \alpha_\tau+\beta_\tau x_i} \tau|y_i - \alpha_\tau - \beta_\tau x_i| \; + \sum_{i:\, y_i < \alpha_\tau+\beta_\tau x_i}(1-\tau)|y_i - \alpha_\tau - \beta_\tau x_i| \qquad (7\text{-}1)$$

where α_τ and β_τ are regression coefficients associated with the quantile τ, ranging between 0 and 1, and $\boldsymbol{y}$ indicates the drought indices (i.e., the SPI and SSI). In this study, the null hypothesis of a zero slope for drought characteristics was tested at a level of 95% at quantile τ.

7.3.2 Overview of Copula function

Because of the interdependence between precipitation and SM, I derive the multivariate standardized drought index (MSDI) by constructing the joint distribution of two drought indices (i.e., the SPI and SSI). Among various types of multivariate models, the copula has been widely applied in various areas including hydrological and climatological applications since the copula can effectively link the marginal distributions together to construct the joint distribution (Favre et al., 2004; Kao and Govindaraju, 2010; Kwon et al., 2016). From a modelling viewpoint, Sklar's Theorem (Sklar, 1959) allows us to model the marginal distributions separately from the dependence structure, which is described by a copula parameter C (Lall et al., 2016; Requena et al., 2013; Rüschendorf, 2009; Salvadori and De Michele, 2004). The proposed approach provides a useful framework to assess overall drought conditions since the MSDI can integrate different aspects of drought dynamics, covering meteorological and agricultural droughts. Here, I briefly present the concept of the copula. For more details, readers are kindly referred to Joe (1997); Nelsen (1999); and Salvadori and De Michele (2004). Let the SPI and the SSI be continuous random variables *X* and *Y*. If a joint distribution exists with the marginal distribution *F*(*X*) and *G*(*Y*), then the cumulative joint probability p with a copula C can be represented as Eq. (7-2).

$$P(X \le x, Y \le y) = C[F(X), G(Y)] = p \quad (7\text{-}2)$$

Finally, the cumulative joint probability p is transformed into the MSDI as follows:

$$\mathrm{MSDI} = \varphi^{-}(p) \quad (7\text{-}3)$$

where φ^{-} is the inverse of the standard normal distribution function. The parameters of the copula functions are estimated using the maximum likelihood (ML) method, and the optimal copula for drought variables (i.e., X and Y) is then selected based on the Akaike Information Criteria (AIC) (Akaike, 1974). Table 7-3 shows copula functions (i.e., Gaussian, t, Clayton, Frank and Cumbel) considered in this study.

Table 7-3 Copula functions considered in this study.

Name	C(u,v)	Reference
Gaussian	$\int_{-\infty}^{\emptyset^{-1}(U)} \int_{-\infty}^{\emptyset^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(\frac{2\theta xy - x^2 - y^2}{2(1-\theta^2)}\right) dxdy$	Li et al., (2013)
t	$\int_{-\infty}^{t_{\theta_2}^{-1}(U)} \int_{-\infty}^{t_{\theta_1}^{-1}(v)} \frac{\Gamma\left(\frac{\theta_2+2}{2}\right)}{\Gamma(\theta_2)\pi\theta_2\sqrt{1-\theta_1^2}} \left(1 + \frac{x^2 - 2\theta_1 xy + y^2}{\theta_2}\right)^{(\theta_2+2)/2} dxdy$	Li et al., (2013)
clayton	$\max\left(U^{-\theta} + v^{-\theta} - 1, 0\right)^{-\frac{1}{\theta}}$	Clayton, (2016)
Frank	$-\frac{1}{\theta}\ln\left[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1}\right]$	Li et al., (2013)
Gumbel	$\exp\left\{-1\left[(-\ln(u))^{\theta} + (-\ln(v))^{\theta}\right]^{\frac{1}{\theta}}\right\}$	Li et al., (2013)

7.3.3 Clustering analysis

The hierarchical agglomerative clustering analysis is employed to classify spatio-temporal regional drought patterns into certain categories. In this manner, weather stations are partitioned into subsets by defining a measure of distance or dissimilarity in terms of drought features. In other words, each category should be mutually exclusive, and the drought characteristics assigned to a certain group should be as similar as possible. The hierarchical agglomerative clustering approach

begins with a measure of the similarity (or dissimilarity) between the objects (i.e., the MSDI time series over 55 weather stations) that are initially regarded as an individual cluster, and the individual clusters are then successively merged until one cluster includes all objects. In this study, Ward's method, which is referred to as an increase of sum-of-squares, is used to assess the proximity between two clusters.

$$d(r,s) = \sqrt{\frac{2n_r n_s}{(n_r+n_s)}}\ \|\overline{x}_r - \overline{x}_s\|_2 \tag{7-4}$$

where $\| \ \|_2$ is the Euclidean distance, $\overline{x}_r$ and $\overline{x}_s$ are the centroids of clusters *r* and *s*, and n_r and n_s are the number of elements in clusters *r* and *s*.

7.4 Results and discussion

7.4.1 Drought identification and relationship between the SPI and SSI

To explore drought propagation, I first evaluate cross-correlations between the SPI and SSI to quantify the lag time over the entire array of weather stations for 3- and 6-month accumulation periods, as shown in Figure 7-2. Here, the strongest cross-correlation at each station is marked by a black dot. It is clearly seen that lag-1 correlation is significant for the majority of the stations, indicating that the meteorological drought on the monthly basis of SPI is identified one month earlier than the SSI. In other words, it may take approximately one month at most weather stations for precipitation deficits to propagate to SM deficits through the hydrological cycle.

In addition to the above drought features, an understanding of drought persistence, which has an impact on water resources management, is also of great interest for hydrologists (AghaKouchak, 2015; Ganguli and Ganguly, 2016; Meng et al., 2017). Drought persistence, like other drought characteristics, varies spatially and temporally (Ford and Labosier, 2014; Meng et al., 2017; Mo and Schemm, 2008), and can be computed by the length of a dry spell for a certain threshold or by using temporal autocorrelations (Tatli, 2015).

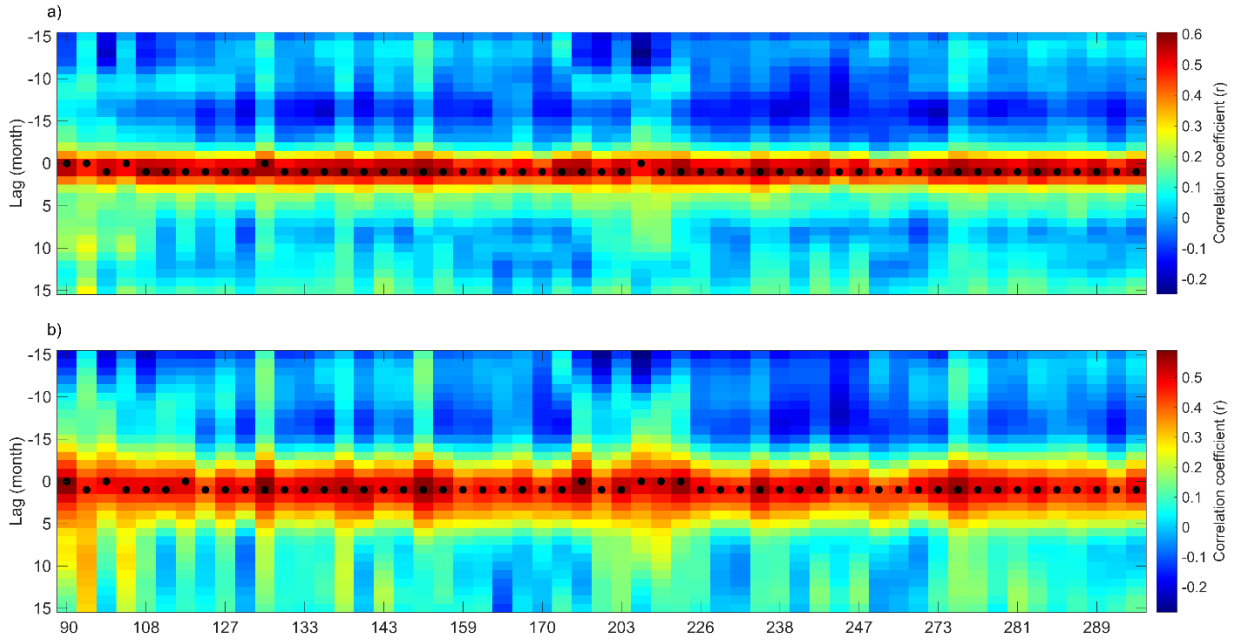


Figure 7-2 Heat maps showing cross-correlation coefficients between SPI and lagged SSI over the entire array of weather stations. Here, a) and b) represent accumulation periods of 3 and 6 months, respectively.

I further explore a monthly variation of the correlation coefficient between the SPI-3 and SSI-3 for the 55 weather stations, and the lagged relationships between the drought indices are additionally examined to capture any possible delayed response. As shown in Figure 7-3, the SPI is in general positively correlated with the SSI, thus confirming that the deficit of SM is substantially related to the meteorological drought (Van Loon, 2015). However, there exist seasonal variations in correlation coefficients, which can be explained by the fact that agricultural drought in response to the deficit in rainfall may differ significantly, depending on the season. Furthermore, it can be concluded that the stronger relationship begins when rainfall starts after a long, dry winter-spring season in South Korea. In this stage, the water moves both through the soil and over the surface via a range of hydrologic processes such as base flow, seepage, infiltration and runoff throughout the summer. In contrast, the relationship weakens as SM content decreases below the wilting point, with the relationship continuing to weaken until the next wet season. Interestingly, it appears that during a dry season (winter-spring), there is a more robust relationship between the SPI and the 1-month lagged SSI. It can be concluded that the SSI has a delayed response to the SPI under dry soil conditions, whereas, for wet soil conditions, the prompt response

of the SSI to the SPI is dominant. In this perspective, it is acknowledged that the characteristics of the transition from meteorological drought to agricultural drought are significantly dependent on antecedent SM content over the season. Thus, we should consider the issue of presenting the role of antecedent SM content in connection with evidence concerning changes in the drought propagation feature over time.

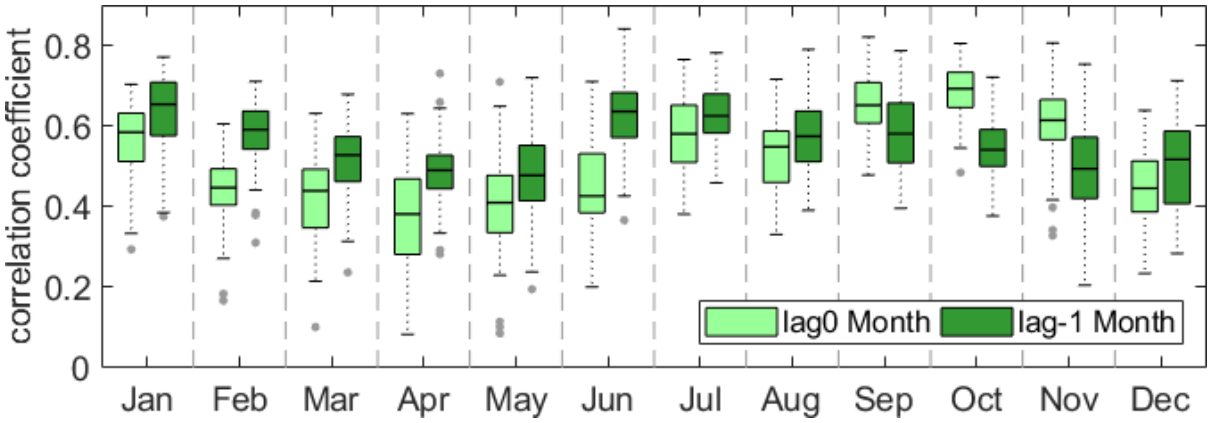


Figure 7-3 Boxplots of the Pearson correlation coefficients for identifying time-lagged relationships between the SPI-3 and SSI-3 time series on a monthly basis across all stations.

I investigate the temporal persistence of the drought indices using the autocorrelation function representing drought persistence over the entire array of weather stations, as illustrated in Figure 7-4. Since the longer-timescale SPI and SSI accumulate observation records over an extended period of time, one can expect that the autocorrelation increases over the accumulation period. As shown in Figure 7-4, it is evident that the autocorrelation functions of SSI decrease gradually with higher degrees of autocorrelation compared to that of the SPI for both accumulation periods. On the other hand, the results also highlight the potential benefit of using different drought indices, i.e., the onset of a drought condition can be detected by the meteorological index (i.e., the SPI) earlier, whereas the SSI seems to be more appropriate for reliably describing drought persistence (Entekhabi et al., 1996; Farahmand and AghaKouchak, 2015). In this regard, I introduced a robust framework that allows multiple drought indices to be combined. The results associated with the combined drought indices are presented in Section 7.4.3.

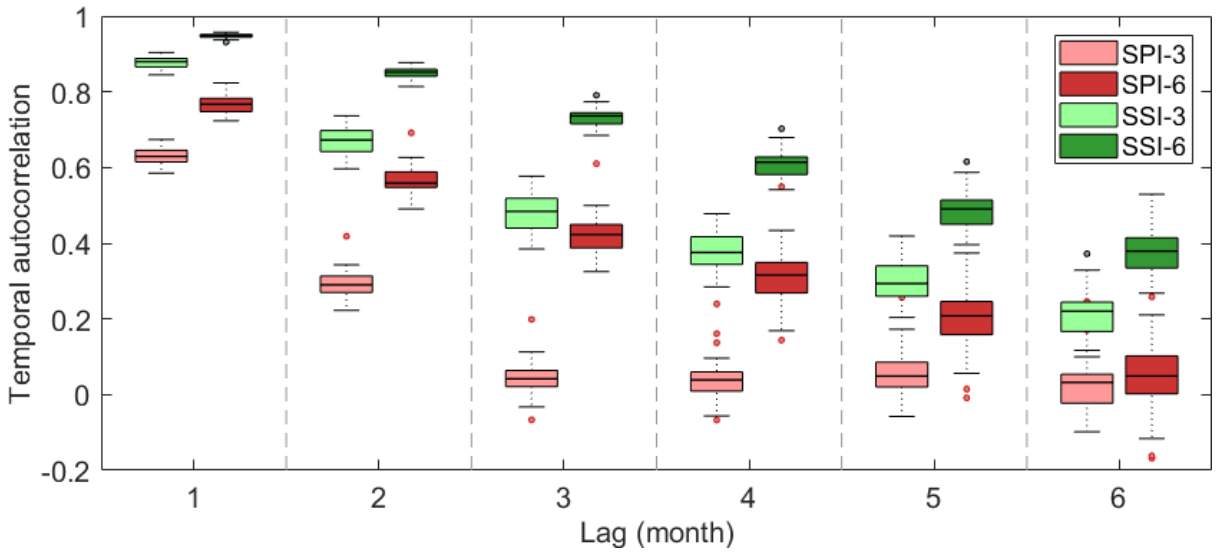


Figure 7-4 Temporal autocorrelation functions of SPIs and SSIs representing drought persistence with respect to different time lags.

7.4.2 Spatial pattern of drought over South Korea

For each accumulation period (3- and 6- month), drought events are identified using a threshold of -1.0 and their spatial distributions along with durations are displayed in Figure 7-5. The kriging spatial interpolation method is hereinafter employed to obtain the regional distribution of drought characteristics. Compared to the SPI, the SSI shows less frequent droughts for both accumulation periods, which can be attributed to the stronger persistence (i.e., a smaller fluctuation) that is more likely to be characterized by the SSI (Farahmand and AghaKouchak, 2015). Furthermore, more frequent drought events appear to occur at a shorter timescale for both drought indices (Figure 7-5c and 7-5g), due to the relatively weaker persistence (Figure 7-5d and 7-5h). It was also clearly seen that drought duration of the SSI is significantly longer than that of the SPI, and the difference becomes more distinct for a longer timescale. In other words, the 3-month drought indices pertain negative values more frequently than do the 6-month drought indices, and the SPI recovers to wet states more quickly than does the SSI. More frequent drought events with the threshold -1.0 particularly stand out in western central South Korea for SPI, while an increased frequency of moderate droughts is identified in the southern and northern parts of South Korea.

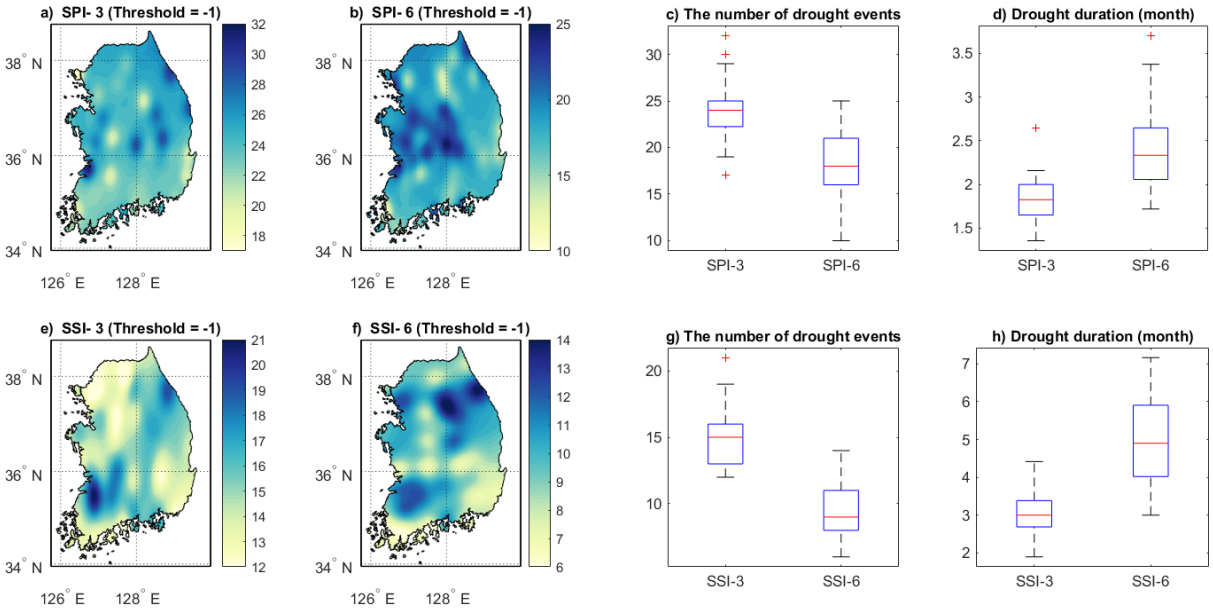


Figure 7-5 Characteristics of SPI/SSI-n drought events based on a threshold of -1.0 (moderate drought): a-b and e-f show the spatial distributions of drought events across South Korea along with their boxplots (c and g), and their corresponding average drought durations are presented in d and h.

Next, drought characteristics such as duration and severity for the SPI and SSI are extracted, and their spatial distributions are presented in Figures 7-6 and 7-7. For SPI-3, a longer drought duration is predominantly identified in the northern and southern parts of South Korea, and the magnitude of drought severity is found to be similar to the spatial distribution of drought duration. Yet, as the accumulation period increases (i.e., SPI-6), the spatial extent of droughts is partially extended to the east coast. As expected, for both drought accumulation periods, the spatial distribution of drought over South Korea clearly reveals the strong spatial coherence between drought duration and severity. That is, regions with a prolonged drought tend to experience more severe droughts, leading to more severe effects on water resource management and vice versa.

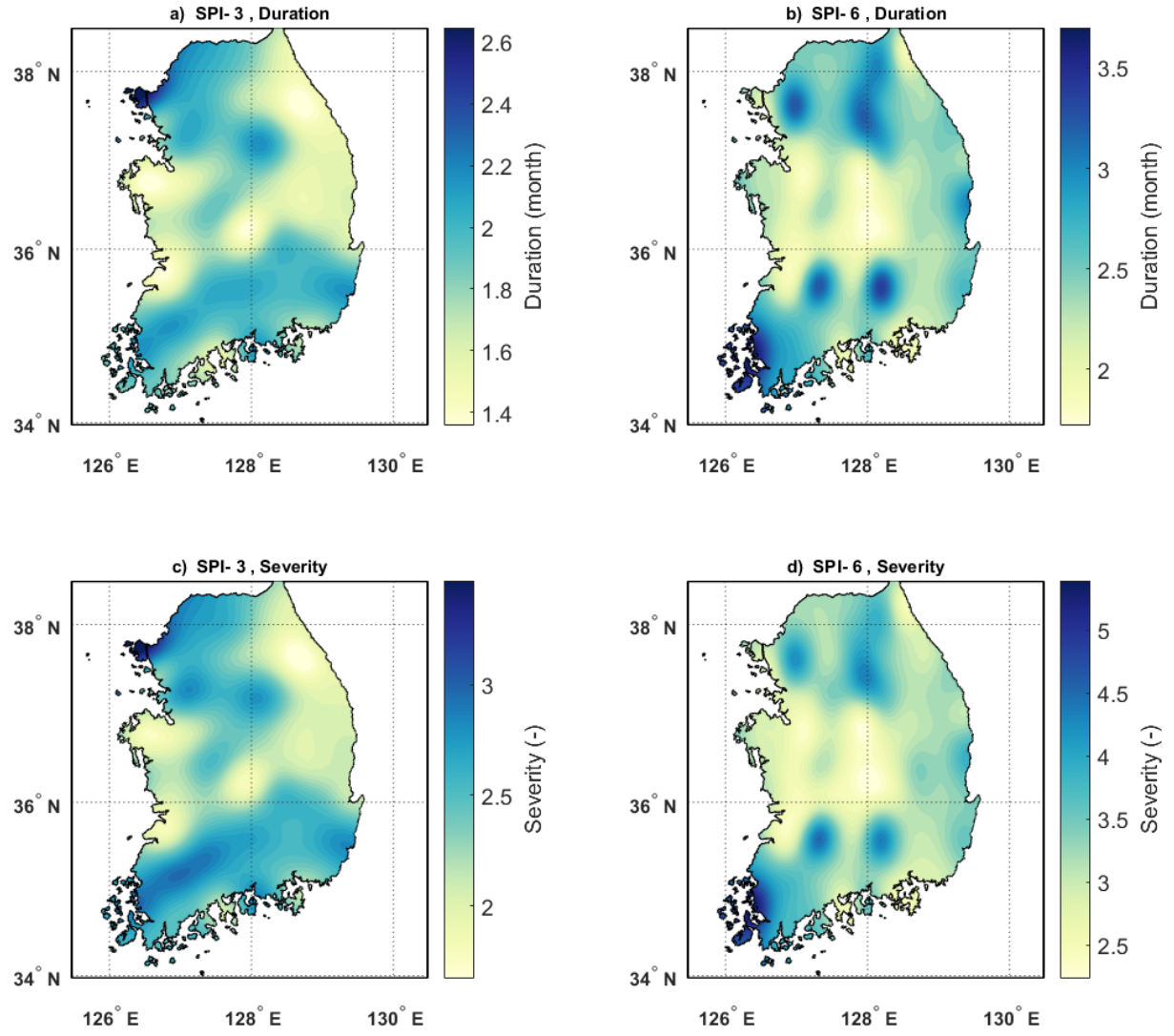


Figure 7-6 Spatial distribution patterns of meteorological drought (SPI-n) duration and severity using a threshold of -1.0 (moderate drought).

Additionally, there is a tendency for duration and severity to increase in proportion to the accumulation period for either SPI or SSI. As shown in Figure 7-7, there are some differences in the spatial distribution of duration and severity between the SSI and SPI. It can be seen that the SSI consistently yields higher drought durations and severities, compared with the SPI. Again, this may be attributed to the stronger persistence of the SSI. Locally significant severe droughts in terms of both drought duration and severity are primarily represented in the southeastern region. Interestingly, the spatial distribution of drought characteristics associated with the SPI is more

dependent on the accumulation periods (see Figure 7-6), while the SSI is less sensitive to the accumulation periods (see Figure 7-7).

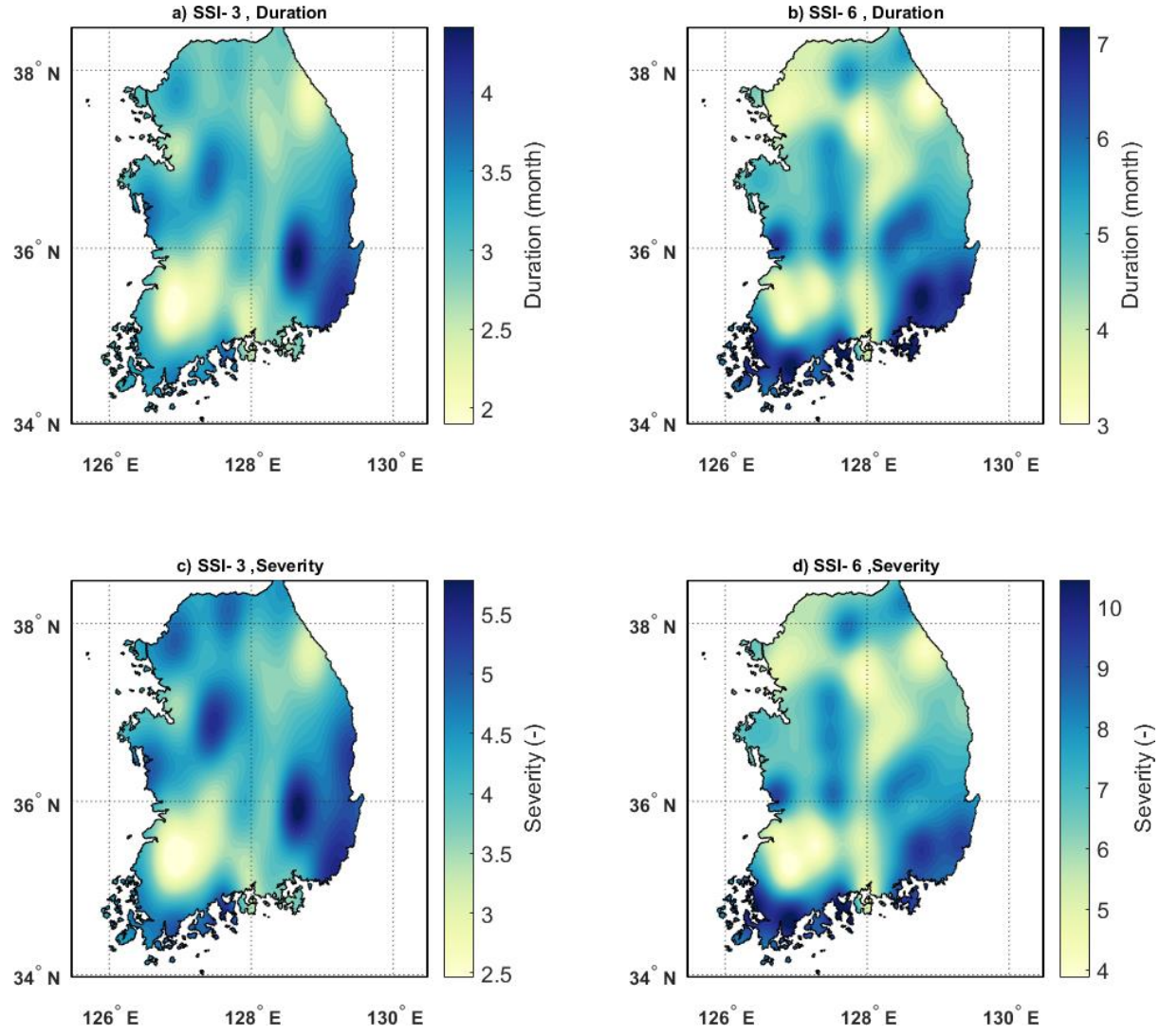


Figure 7-7 Spatial distribution patterns of agricultural drought (SSI-n) duration and severity using a threshold of -1.0 (moderate drought).

The trends in the SPI and SSI time series covering 1986-2016 are analyzed using a quantile regression model. The quantile-based regression model is employed to identify temporal variations at different quantile levels. In particular, quantile regression analysis was performed on the SPI and SSI time series using the predefined drought categories (i.e., thresholds) represented in Table 7-2. The estimated slope parameters for the three different thresholds at 55 weather stations are spatially interpolated and illustrated in Figure 7-8. For the sake of brevity, I only provide the results

based on the 3-month accumulation period. As shown in Figures 7-8a and 7-8d, for both indices, the trends for moderate drought (i.e., threshold -1.0) showed a downward tendency in the northern part of South Korea, while an upward trend is dominantly localized in the southern region. This distribution has a similar spatial structure to that of severe drought (i.e., threshold -1.5), as shown in the SSI-3. Contrary to the SSI, upward trends in the SPI are spread over the entire central region for severe drought, while a decreasing trend appears over the northern part of South Korea. There exist significant differences in the spatial presence of trends in extreme drought (i.e., threshold -2.0) between the SPI and SSI. More specifically, the SSI shows a decreasing tendency over the entire region, while there is no significant difference in the spatial distribution of drought trend over different drought states for the SPI. In summary, there appears to be a more pronounced decreasing tendency (or increasing risk) of the drought in the northern part of South Korea. The increased drought risk in the northern part of South Korea may be related to a weakened tendency of summer monsoons over the last three decades, which is mainly associated with stagnation of the monsoon belt in the middle of South Korea (Zhang and Zhou, 2015).

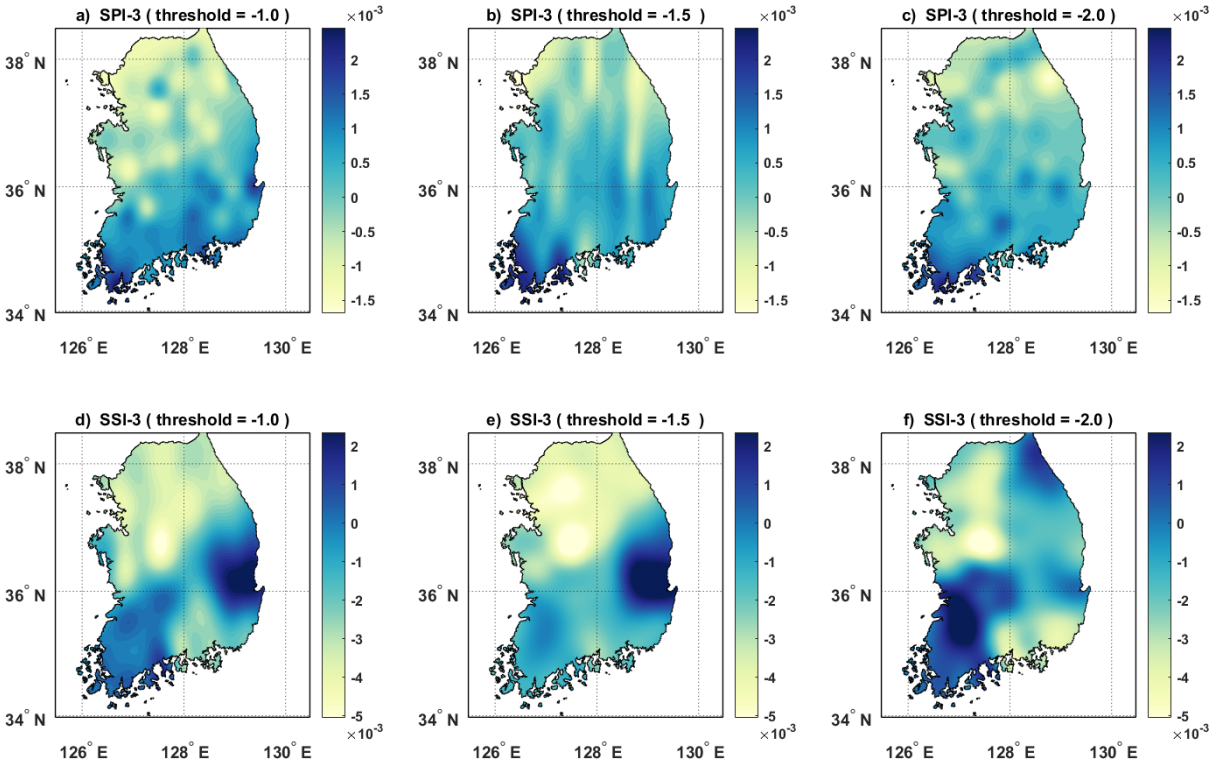


Figure 7-8 Spatial distribution of trends in SPI-3 and SSI-3 at different quantile levels.

7.4.3 Clustering analysis on Multivariate Standardized Drought Index (MSDI)

A hierarchical clustering approach is applied to explore regional trends in droughts over the last three decades. One may consider a direct use of two drought indices for clustering regional patterns. However, based on the preliminary analysis, the distribution of the identified clusters in the SPI and SSI are significantly different, thus confirming that the direct use of indices together for the clustering may fail to identify regional patterns of drought. On the other hand, with multi-dimensional data, one can employ multivariate techniques, such as copulas, which can provide a better estimation for dependencies among the variables, prior to clustering. Specifically, I introduce the MSDI to offer a comprehensive perspective of the drought by constructing a joint probability distribution between the SPI and SSI. I consider two types of elliptical copulas and three types of Archimedean copulas to model the dependency structure of the drought indices, namely the student t, Gaussian, Clayton, Frank and Gumbel. As discussed in the methodology section, the marginal distributions are first specified with the independent identically distributed (IID) assumption for the drought variables, and the interdependence between the variables is then described through the copula functions. In this study, it might be desirable to assume the Gaussian distribution for the marginal distributions for both the SSI and SPI, since the indices are already normalised to their respective values. A set of parameters for the MSDI, four parameters in the marginal distribution and one parameter in the copula function, are estimated by using the maximum likelihood method (Bouyé et al., 2000; Kao and Govindaraju, 2010; Renard and Lang, 2007), and the optimal copula functions are then selected using the Akaike Information Criteria (AIC) for each weather station. Among five types of copulas, the Frank copula is generally chosen for both accumulation periods, as summarised in Table 7-4.

I use clustering analysis to explore the presence of a regional trend in drought, and an important issue with respect to the clustering approach is to determine the number of desired clusters. To systematically choose the optimum number of clusters with the hierarchical agglomerative clustering algorithm, the algorithm is recursively applied to the MSDI series with an increasing number of clusters and the optimum number of clusters is selected by maximising (or minimising) some measure of fitness. This study uses the upper-tail rule, proposed by Mojena (1977), as a measure of model fitness. The best cut-off level (i.e., the number of clusters) is determined by the distance analysis of the standardized fusion levels in a dendrogram. As shown in Figure 7-9, the

inflexion point of the MSDI-3 is found at a cut-off level of four (Figure 7-9a). In other words, the degree of decrease in the standardized fusion level is negligible for more than four clusters. On the other hand, an inflexion point is found at five clusters for the MSDI-6 (Figure 7-9b). Therefore, four clusters for the MSDI-3 and five clusters for the MSDI-6 are subsequently selected for further analyses.

Table 7-4 Summary of the copula family for each weather station along with the corresponding cluster.

MSDI-3						MSDI-6					
Sta. No.	**Copula**	**Cluster No.**	**Sta. No.**	**Copula**	**Cluster No.**	**Sta. No.**	**Copula**	**Cluster No.**	**Sta. No.**	**Copula**	**Cluster No.**
90	Frank	3	203	Frank	4	90	Gaussian	1	203	t	2
100	t	3	211	Frank	4	100	Frank	1	211	Gaussian	1
101	Gumbel	3	212	Gaussian	1	101	Gaussian	1	212	Clayton	1
105	t	3	221	Gaussian	1	105	Frank	1	221	Frank	2
108	Gumbel	3	226	Frank	4	108	Gumbel	2	226	Frank	5
112	Gumbel	3	232	Frank	4	112	Gumbel	2	232	Frank	5
114	Gaussian	4	235	Frank	4	114	Gaussian	2	235	Frank	5
119	Frank	3	236	Frank	4	119	t	2	236	Frank	5
127	Frank	4	238	Frank	4	127	Frank	2	238	Frank	5
129	Frank	4	243	Frank	4	129	Frank	5	243	Frank	5
130	t	1	244	Frank	4	130	t	3	244	Gaussian	5
131	Frank	4	245	Frank	4	131	Frank	5	245	Gaussian	5
133	Frank	4	247	Gaussian	1	133	Frank	5	247	Gaussian	5
135	Frank	1	260	Gumbel	5	135	Frank	3	260	Frank	4
138	t	1	262	Gumbel	5	138	t	4	262	Gaussian	4
140	Frank	4	272	Gumbel	5	140	Frank	5	272	Frank	3
143	Gaussian	2	273	Frank	4	143	Frank	4	273	Frank	3
146	Frank	4	277	t	2	146	Frank	5	277	t	3
152	Frank	2	278	Gaussian	1	152	Frank	4	278	Frank	3
156	Frank	4	279	Frank	4	156	Frank	5	279	Frank	3
159	Gumbel	2	281	Gaussian	1	159	t	4	281	t	4
162	Gumbel	2	284	t	2	162	Frank	4	284	Frank	4
165	Gumbel	2	285	Gaussian	1	165	Frank	4	285	Frank	4
168	Gaussian	2	288	Gumbel	5	168	Frank	4	288	Frank	4
170	Gumbel	2	289	Frank	4	170	Frank	4	289	Frank	4
192	Frank	2	294	Gumbel	5	192	Frank	4	294	Frank	4
201	Gumbel	3	295	Gumbel	5	201	Gumbel	2	295	Frank	4
202	Gaussian	3				202	Clayton	2			

The distribution of the resulting clusters is contiguous rather than spatially separated for both accumulation periods, as presented in Figure 7-10. It can be concluded that the results are more physically interpretable, which can lead to more effective strategies in creating drought mitigation plans for certain areas. There is a notable contrast to the clustering over accumulation periods in the northern part of South Korea, namely Gyeonggi province and Gangwon province. More specifically, two subcategories, representing Gyeonggi and Gangwon province, in MSDI-6 are grouped together as a category in MSDI-3. For a given cluster, this study further explores different aspects of drought features such as duration, severity and long term trends at different quantile levels.

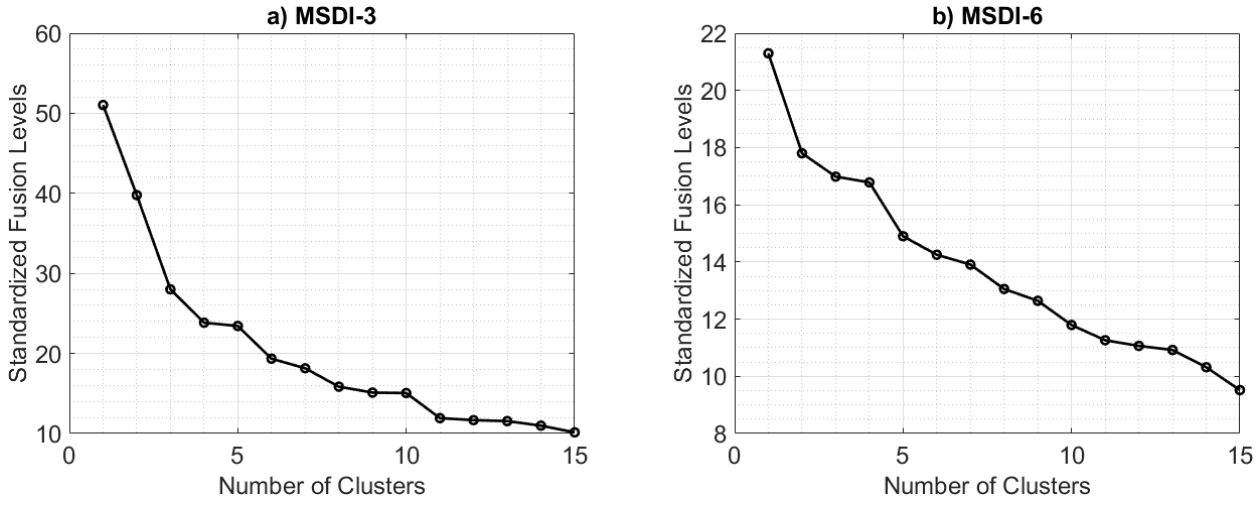


Figure 7-9 Standardized fusion levels corresponding to the number of clusters for MSDI-3 and MSDI-6.

The spatially averaged MSDI values over each cluster are shown in Figure 7-11, and their trends appear to differ significantly between thresholds (or exceedance probabilities; moderate (-1.0): 0.16, severe (-1.5): 0.07 and extreme (-2.0): 0.02). Compared to the MSDI-6, more frequent drought events appear to be identified at the MSDI-3, and the number of drought events for certain drought categories varies over different clusters, as summarised in Table 7-5. It is evident that drought duration of the MSDI-6 is significantly longer than that of the MSDI-3, and the difference becomes more distinct under extreme drought conditions. Similarly, an overall increase in drought severity in the MSDI-6 is clearly observed. Specifically, more moderate to severe drought episodes are observed in the northern part of South Korea, covered by clusters CL-3 and CL-4 for the MSDI-

3 and by clusters CL-1 and CL-2 for the MSDI-6. Clusters CL-2 for the MSDI-3 and CL-4 for the MSDI-6 notably for the longer duration period and the higher severity are mainly identified along the southern coast, under moderate to severe drought condition. Spatially aggregated drought features vary under extreme drought conditions. In other words, CL-3 for the MSDI-3 indicates a much more extreme drought condition in the northern part of South Korea, and as does CL-4 for the MSDI-6 in the southern coast.

Additionally, I explore regional trends over predefined thresholds using a quantile regression model, as represented in Figure 7-11 and Table 7-6. Numbers in bold are statistically significant at the 0.05 level ($p<0.05$) in Table 7-6. For the MSDI-3, CL-1 representing drought in the central eastern region showed no trend for all four levels, while a significantly decreasing trend is shown in CL-3. A significantly decreasing trend (or increasing drought risk) in MSDI-3 was found at CL-2 (e.g., median) and CL-4 (e.g., extreme, moderate and median). Overall, for a longer-duration MSDI-6, a significant downward trend becomes more dominant (see Figure 7-11(e) -711(i). Furthermore, I explore the past three major drought episodes over the last three decades (1986-2016), as shown in Figure 7-11. As presented, drought episodes Ep1 (1994-1996), Ep2 (2000-2002) and Ep3 (2013-2015) are clearly identified as major drought events that have been reported in previous studies (Kwon et al., 2016; Min et al., 2003; Nam et al., 2015).

Table 7-5 Summary of drought episodes based on clustering analysis.

Threshold	The number of Events					Duration					Severity				
	CL-1	CL-2	CL-3	CL-4	CL-5	CL-1	CL-2	CL-3	CL-4	CL-5	CL-1	CL-2	CL-3	CL-4	CL-5
MSDI-3															
Moderate	33	29	**41**	34		3.4	**4.5**	3.0	3.5		5.5	**7.2**	4.6	5.6	
Severe	22	25	**26**	**26**		2.6	**2.8**	2.2	2.5		5.1	**5.4**	4.2	4.8	
Extreme	11	**13**	7	11		1.8	1.6	**2.3**	1.6		4.4	3.8	**5.8**	3.9	
MSDI-6															
Moderate	**25**	**25**	23	20	24	4.6	4.8	5.2	**6.4**	5.0	7.1	7.6	8.3	**10.3**	8.2
Severe	17	**20**	16	17	19	2.6	2.7	**3.9**	3.8	3.8	5.4	5.3	**7.6**	7.4	7.4
Extreme	7	7	**11**	9	12	3.1	2.7	2.3	**3.3**	2.0	7.5	6.6	5.3	**7.5**	4.7

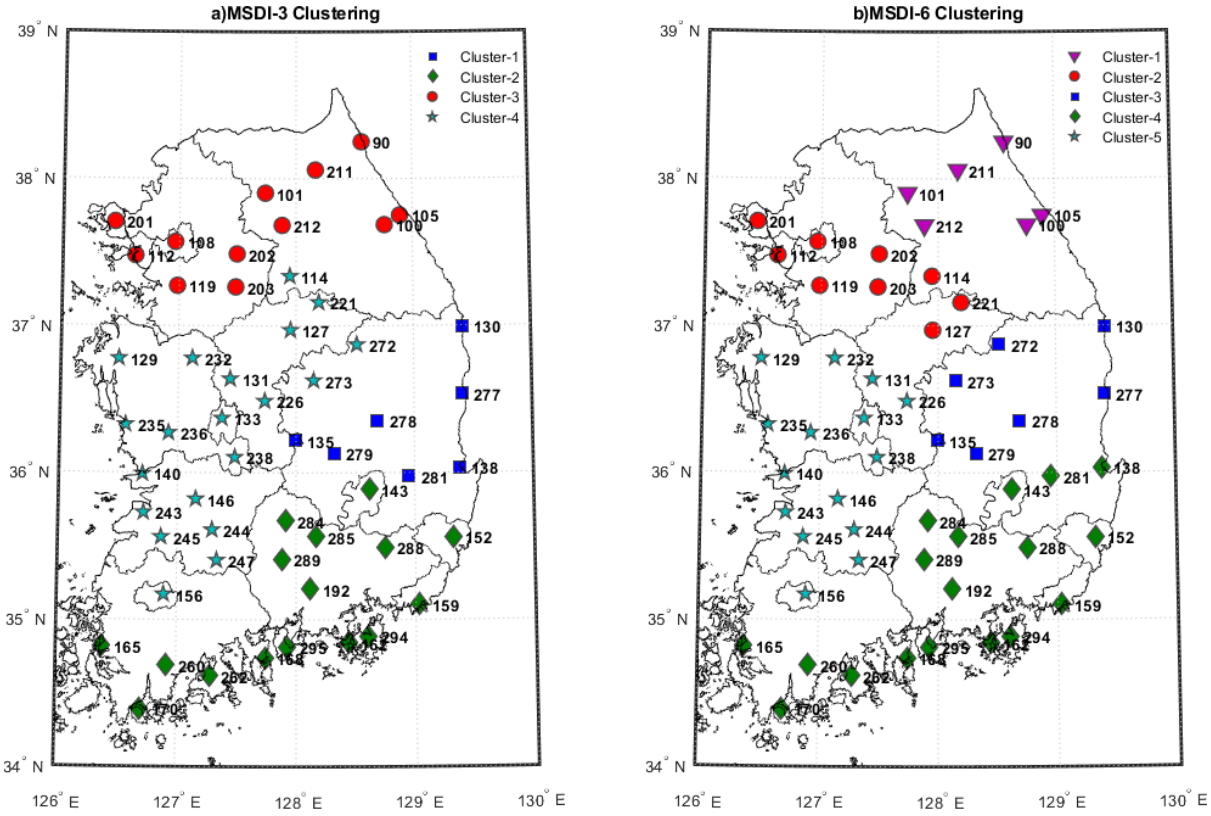


Figure 7-10 Spatial distribution of clusters for the MSDI over South Korea. Left panel, MSDI-3; right panel, MSDI-6.

Table 7-6 Summary of the slope obtained from a quantile regression model at four different classes. Numbers in bold are statistically significant at the 0.05 level.

Cluster	MSDI-3 (× 10^{-3})				MSDI-6 (× 10^{-3})			
	Extreme	Severe	Moderate	Q50	Extreme	Severe	Moderate	Q50
CL-1	1.14	0.07	-0.79	-1.21	**-3.55**	**-3.17**	**-1.60**	**-2.00**
CL-2	1.13	-0.34	-0.07	**-1.39**	**-3.63**	**-3.07**	**-3.19**	**-1.82**
CL-3	**-3.51**	**-3.29**	**-1.79**	**-1.78**	-1.36	-1.15	-1.11	**-1.96**
CL-4	**-2.21**	-0.77	**-1.70**	**-2.28**	0.09	-0.13	-1.10	**-1.92**
CL-5					-2.69	-1.49	-1.32	**-2.37**

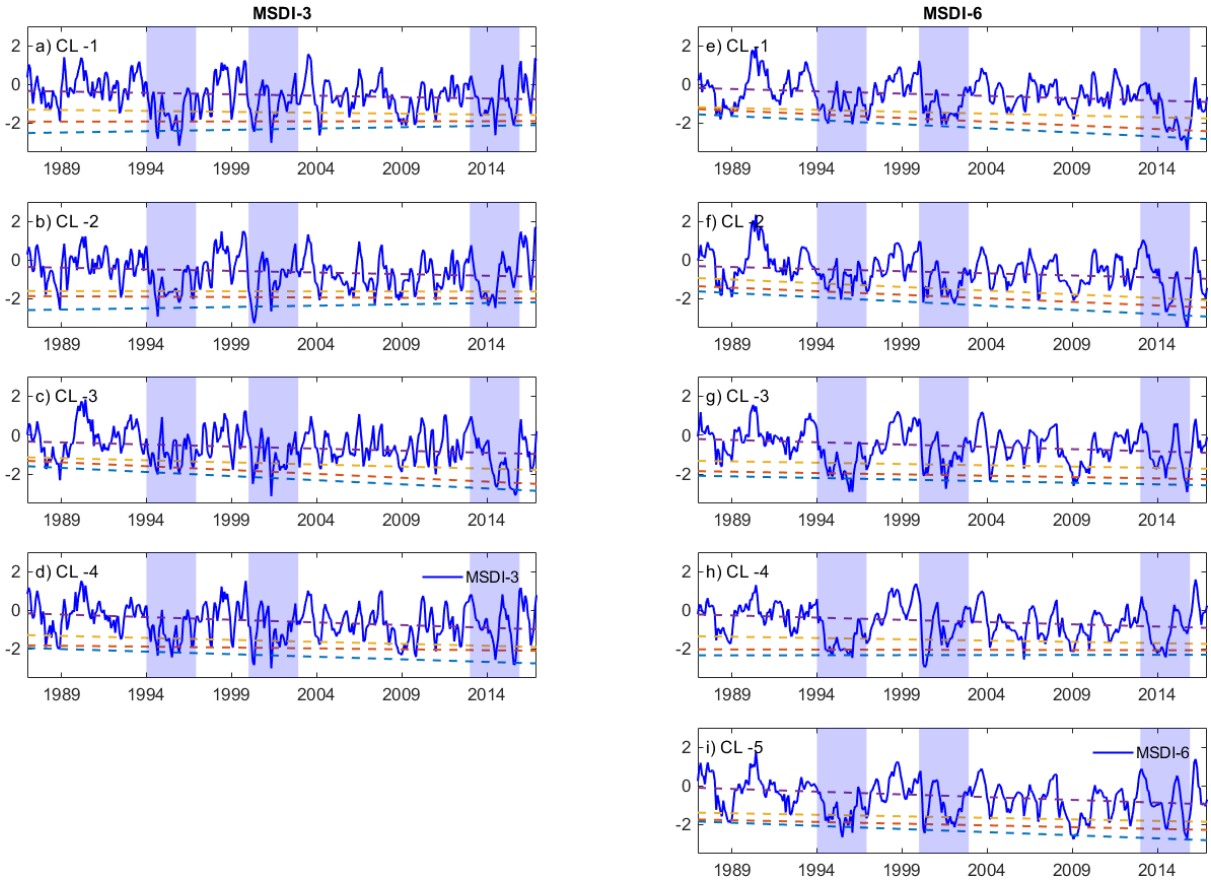


Figure 7-11 Regional trends of MSDI-3 and MSDI-6 corresponding to each cluster and their trends over different thresholds based on quantile regression. Here, blue bars denote three major drought episodes over the past three decades.

7.5 Conclusions

Drought is an increasingly important issue in many parts of the world, requiring a hydro-meteorological modelling framework to assess and monitor its complex impact on natural hazards and associated socio-economic vulnerability. In this chapter, I use two representative drought indices (the SPI and SSI) to evaluate changes in drought patterns at different spatio-temporal scales. The SPI and SSI, derived from precipitation and SM, respectively, are compared with each other by describing their individual characteristics as drought indicators as well as their interdependence and interaction. Furthermore, considering different aspects of the drought dynamics, this chapter introduces the MSDI, which is used to explore meteorological and agricultural droughts jointly in the context of a multivariate probability distribution. The MSDI derived for each station is then

grouped using the hierarchical clustering approach for better understanding of the regional features of drought conditions. The primary conclusions obtained in this chapter are as follows:

(1) The transition from meteorological to agricultural drought is clearly identified, but the degree of their relationship is significantly dependent on the season. Specifically, the SSI had a 1-month delayed response to the SPI during the dry season (i.e., winter-spring), whereas the response of the SSI to SPI is generally prompt under wet soil conditions. Thus, one should consider the role of antecedent SM content to improve the characterization of changes in drought propagation.

(2) The SSI shows less frequent droughts and longer drought duration, due to the gradual decrease in the autocorrelation functions of SSI along with the higher degree of autocorrelation, compared to that of the SPI. In this perspective, the onset of drought could be detected by the SPI, whereas the SSI appears to be more appropriate for describing drought persistence. Overall, this is also supported by the fact that the 1-month lag between the SPI and SSI was significant for most stations over the last three decades (1986-2016).

(3) The copula-based MSDI is employed to consider the interdependence and interaction between rainfall and SM in the context of a multivariate probability distribution. Moreover, the hierarchical agglomerative clustering approach is used to identify the spatial pattern of the MSDI. The distribution of the resulting clusters is contiguous rather than spatially isolated for both accumulation periods, contributing to more effective strategies in the development and implementation of drought management and mitigation plans for certain areas.

(4) In addition, I use a hierarchical clustering approach to the MSDI to investigate regional trends in drought pattern. By using this approach, the spatio-temporal drought patterns are clearly captured through the MSDI. Specifically, more drought episodes under moderate to severe drought conditions are dominantly observed along the southern coast of South Korea. I also find persistent drought with a higher level of severity in the northern part of South Korea, which might be attributable to the significant decreasing trend (or increasing drought risk) that is noted in the northern part of South Korea. Overall, for a longer-duration MSDI-6, a significant downward trend has become more dominant.

The main contribution in this chapter is to statistically combine the SPI with the SSI with the intention of providing a more comprehensive perspective for drought characterisation. Although

the case study site is in South Korea, this chapter is expected to provide valuable insights into the enhanced applicability of SM information in drought assessment for decision-making.

CHAPTER 8 Conclusions and recommendations

8.1 Conclusions

This thesis aims to explore the particular features of multiple SM data sources, in which each dataset tends to show different advantages and limitations in terms of practical aspects. Moreover, given that SM knowledge could provide a more comprehensive perspective of RR simulation and drought assessment, SM data are incorporated into the hydrological modelling framework and further employed to identify the characteristics of historical droughts in South Korea. This thesis can be partitioned into two main parts as stated in Chapter 1.

The first part (Chapter 4 and Chapter 5) concentrates on assessing microwave SM products and addressing their inherent limitations, such as a shallow observation depth and a coarse spatial resolution. The main findings of the first part are as follows:

(1) The RZSM derived from the exponential filter method successfully reproduces SM content in a deeper layer, providing improvements in accuracy compared to the original SM. Satellite-retrieved SM is representative of a topsoil layer, while the RZSM is more readily applicable to be incorporated into hydro-meteorological models. In this sense, this approach used in this chapter can be directly employed in various hydrological applications. For instance, prior to integrating satellite SM into a hydrological model, this approach is used as a pre-processing stage in Chapter 6.

(2) Aside from the filtering method, the CDF matching method is applied, not only to address the inherent systematic mismatch between the estimated RZSM and in-situ SM but also to select an optimal temporal resolution. As a result, compared to the conventional CDF matching method that uses the whole record of the investigation period, a bias-correction period of QM4 (growing and non-growing season) outperforms the other temporal groups for both calibration and validation periods. It should be noted that the results achieved in this thesis might be location-dependent, meaning that different optimal temporal group can be obtained from other locations. However,

given that little work on this topic has been conducted to explore the optimal bias-correction period in the literature, the methodology and results of the current study are useful in the field of remote sensing and relevant to the wider hydrological community.

(3) The efficiency of the proposed GM-NHMM method is compared with that of the OLR model using the same predictors. The mean correlation coefficient of the proposed model equal to 0.78, which is significantly greater than that of the OLR model (0.49). Moreover, the proposed model in this thesis preserves the spatial coherence across all stations reasonably well (Figure 5-11), which is a fundamentally important property in describing the spatial pattern of SM and its association with runoff on a catchment scale.

(4) Rainfall and air temperature are considered predictors in the proposed downscaling modelling framework due to the ease of data acquisition and their interaction with SM state in the hydrological cycle. These variables' correlations with in-situ SM are statistically significant and strongly persistent. The lag-1 correlation is high for daily rainfall, and the correlations appear to be consistent with the lag in the temperature and AMSR2 data. It is therefore recommended that researchers should take the time lag relationship between in-situ SM observations and the candidate predictors into account in the proposed GM-NHMM model.

(5) It is clearly seen that the rainfall predictor plays a substantial role in achieving the overall predictability. Using weather variables (i.e., rainfall and temperature) can be effective in picking up some of the predictability of local SM that is not captured by the AMSR2 data. On the other hand, larger-scale dynamic features identified from the AMSR2 data seem to facilitate the identification of regional spatial patterns of SM.

The second part of this thesis (Chapter 6 and Chapter 7) is dedicated to exploring the potential of SM in the context of hydrological applications such as RR modelling and drought assessment. Chapter 6 focuses on improving the ability of RR modelling by combining satellite SM products with an existing RR model (the Tank model) on the basis of a machine learning framework (i.e., the LSSVM). Further, considering that SM is a crucial hydrological variable for drought assessment, Chapter 7 explores the applicability of SM information for a better understanding of the characteristics of droughts. The main findings of the second part are as follows.

(6) As for the LSSVM-based RR model that uses two lagged values of predictors (i.e., rainfall and ESA CCI SM), the results reveal that satellite SM plays an efficient role in the proposed RR modelling framework. In terms of NSE, the proposed model shows "very good" performance, depending on different combinations of the time-lagged input variables. The results also support the potential benefits of introducing the satellite-based SM estimates into hydrological applications.

(7) The SM state variables derived from the Tank model are successfully incorporated into a LSSVM based regression framework. The enhanced performance of the Tank-LSSVM over the Tank model is particularly found in the simulation of low flows whose accuracy significantly influences a key aspect in the water resources management and planning. However, the contribution of satellite SM data as an external source to runoff simulation in the Tank-LSSVM model is found to be insignificant. This confirms that SM state variables derived from the well-calibrated continuous RR model better represents SM dynamics than those obtained from satellite SM data.

(8) In Chapter 7, the SSI shows less frequent droughts and longer drought duration compared to the SPI, due to the gradual decrease in the autocorrelation functions of the SSI. The strongest cross-correlations are observed at a 1-month lag between the SPI and SSI for most stations. Thus, the SPI is more appropriate for defining the onset of a drought, whereas the SSI appears to be more effective for describing drought persistence. Moreover, the transition from meteorological to agricultural droughts significantly depends on the season, indicating that the transition between them is highly correlated with antecedent moisture conditions.

(9) The copula-based MSDI is introduced to explicitly postulate interdependence between the SPI and SSI in the context of a multivariate probability distribution, and then a hierarchical agglomerative clustering approach along with a quantile regression model is employed to explore regional trends in drought. More drought episodes under moderate to severe drought conditions are observed along the southern coast of South Korea, while persistent droughts with higher severity are observed in the northern part of South Korea.

(10) The two drought indices introduced in this thesis show their particular advantage of assessing the nature of drought. However, considering that drought is a multidimensional phenomenon

and using a drought index based on a single indicator is insufficient in describing the complex aspects of drought, the use of the ERA-Interim reanalysis SM data in drought analysis is expected to provide useful guidelines to detect the nature of drought for regions suffering from a lack of in-situ SM observations.

8.2 Recommendations for future work

The strengths and limitations of the multi-source SM products are highlighted, and this thesis further enhances the utilisation of the SM data in the context of hydrological applications. However, there still remain future works to tackle this thesis's shortcomings, as listed below.

(1) Despite the primary contributions of this study, as for the proposed CDF matching method in Chapter 4, more stable and comprehensive results are expected with a more extended record period. Moreover, because the proposed bias correction method has been primarily carried out in South Korea, it is necessary to apply this approach to different parts of the world to overcome regional perspectives.

(2) For a characteristic time (T), it is found that a smaller T value is obtained in the deeper soil layer at some observation sites, which is not in accordance with the fundamental assumptions of the exponential filter method. In this context, further works are required to explore unknown factors that may also influence T values, namely the proposed scheme should be tested over a wide range of vegetation, soil and climate conditions.

(3) The relatively low revisit frequency of satellite sensors (i.e., low temporal resolution (1-5 day revisit time)) is another practical issue in hydrological perspective. In addition to a spatial downscaling approach, the downscaling scheme proposed in Chapter 5 could provide a useful tool for a temporal downscaling purpose. Further, other remote sensing products (e.g., MODIS products) with fine-scale resolution can be used as auxiliary information with the intention of obtaining higher spatial resolution in future work.

(4) In Chapter 6, given an encouraging improvement in runoff simulation with the proposed hybrid RR model, the modelling framework of the current study could be beneficial and relevant to a

number of different applications in the field of hydrology. However, it is necessary for future work to explore the applicability of either other conceptual models or physically-based models for different regions with a longer period of record, which is required to support the findings of this study.

(5) In addition, it should be noted that combining existing RR models with an LSSVM-based RR framework outperforms models that use satellite SM data. This is partially because the lumped model used in this thesis use the mean value over the catchment area rather than spatially distributed SM. This can lead to a misrepresentation of overall performance in the proposed modelling framework. Future works will be dedicated to addressing this issue.

(6) Understanding the uncertainty in the hydrological model allows hydrologists to gain useful insights to address the shortcoming of the performance in RR simulation. However, little work has been carried out in the context of data-driven RR modelling frameworks. In this regard, another interesting study is to estimate uncertainty in the proposed RR simulation made by the Tank-LSSVM model, which will analytically provide an effective manner to improve model performances.

(7) In Chapter 7, integration with other drought indicators (e.g., streamflow and groundwater) is still required for a more comprehensive understanding of multi-dimensional aspects of drought in a future study. In addition, future work should focus on the extension of data records with multiple predictors such as satellite remote sensing data and hydro-meteorological variables as a potential predictor.

(8) This study explores the composite features of droughts from a historical perspective (i.e., over the last three decades). However, their frequency and severity may differ in coming decades due to the impact of climate change. Therefore, it is necessary for appropriate adaptation and mitigation actions to determine how climate change and global warming affect the nature of drought hazard, which could be conducted by coupling climate models.

(9) As stated in Chapter 3, two-thirds of the annual rainfall occurs during the summer monsoon season, which is the primary source of water resources in South Korea. It has been reported that a general declining trend is evident in the East Asian summer monsoon and this tendency appears to be more prevalent since the early 1990s (Li et al., 2017). In this context, it is

necessary to further explore the relationship between the results in this thesis (i.e., increased drought risk) and weakened summer monsoons.

APPENDIX

Appendix A.

The Tank models with vertically interconnected multiple tanks (three tanks for this study) simulate RR processes such as flood events and continuous runoff. In terms of the 3-Tank model, the side outlets in the first tank represent surface runoff (q_{12} and q_{11}), while the outlets in the second and third tank are considered as intermediate runoff (q_2) and base flow (q_3) respectively. The outputs of each side outlet are calculated based on following formulas:

1st tank

$$\frac{dST_1}{dt} = RF - ET - q_{12} - q_{11} - I_1 \quad \text{(A1)}$$

$$if\ ST_1 > H_{12},\ q_{12} = (ST_1 - H_{12}) \times a_{12};\ if\ ST_1 \leq H_{12}, q_{12} = 0 \quad \text{(A2)}$$

$$if\ ST_1 > H_{11},\ q_{11} = (ST_1 - H_{11}) \times a_{11};\ if\ ST_1 \leq H_{11}, q_{12} = 0 \quad \text{(A3)}$$

$$I_1 = ST_1 \times b_1 \quad \text{(A4)}$$

2nd tank

$$\frac{dST_2}{dt} = I_1 - I_2 - q_2 \quad \text{(A5)}$$

$$if\ ST_2 > H_2,\ q_2 = (ST_2 - H_2) \times a_2;\ if\ ST_2 \leq H_2, q_2 = 0 \quad \text{(A6)}$$

$$I_2 = ST_2 \times b_2 \quad \text{(A7)}$$

3rd tank

$$\frac{dST_3}{dt} = I_2 - q_3 \quad \text{(A8)}$$

$$q_3 = ST_3 \times a_3 \quad \text{(A9)}$$

Total runoff

$$Q = q_{11} + q_{12} + q_2 + q_3 \quad \text{(A10)}$$

where $\frac{dS}{t}$ is the time-dependent tank storage. P, ET and I refer to rainfall, evapotranspiration and infiltration, respectively.

REFERENCES

AghaKouchak, A., 2015. A multivariate approach for persistence-based drought prediction: Application to the 2010-2011 East Africa drought. J. Hydrol. 526, 127–135. https://doi.org/10.1016/j.jhydrol.2014.09.063

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Trans. Automat. Contr. 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Albergel, C., de Rosnay, P., Balsamo, G., Isaksen, L., Muñoz-Sabater, J., 2012. Soil Moisture Analyses at ECMWF: Evaluation Using Global Ground-Based In Situ Observations. J. Hydrometeorol. 13, 1442–1460. https://doi.org/10.1175/JHM-D-11-0107.1

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., Hasenauer, S., 2008a. An evaluation of ASCAT surface soil moisture products with in-situ observations in southwestern France. Hydrol. Earth Syst. Sci. Discuss. 5, 2221–2250. https://doi.org/10.5194/hessd-5-2221-2008

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, a., Piguet, B., Martin, E., 2008b. From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. Hydrol. Earth Syst. Sci. Discuss. 5, 1603–1640. https://doi.org/10.5194/hessd-5-1603-2008

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., Ab, W., 1998. Crop evapotranspiration - Guidelines for computing reference crop evapotranspiration 1–15. https://doi.org/10.1016/j.eja.2010.12.001

Andreasen, M., Jensen, K.H., Desilets, D., Franz, T.E., Zreda, M., Bogena, H.R., Looms, M.C., 2017. Status and Perspectives on the Cosmic-Ray Neutron Method for Soil Moisture Estimation and Other Environmental Science Applications. Vadose Zo. J. 16, 0. https://doi.org/10.2136/vzj2017.04.0086

ASCE, 1993. Criteria for evaluation of watershed models. J. Irrig. Drain. Eng. 119, 429–442.

Barrett, B., Petropoulos, G., 2013. Satellite Remote Sensing of Surface Soil Moisture. Remote Sens. Energy Fluxes Soil Moisture Content 85–120. https://doi.org/doi:10.1201/b15610-6

Basri, H., 2013. Development of Rainfall-runoff Model Using Tank Model: Problems and Challenges in Province of Aceh, Indonesia. Aceh Int. J. Sci. Technol. 2, 26–36. https://doi.org/10.13170/aijst.2.1.574

Behzad, M., Asghari, K., Eazi, M., Palhang, M., 2009. Generalization performance of support vector machines and neural networks in runoff modeling. Expert Syst. Appl. 36, 7624–7629. https://doi.org/10.1016/j.eswa.2008.09.053

Bell, J.P., 1968. Soil moisture estimation by the neutron scattering method in Britain. J. Terramechanics 5, 65. https://doi.org/10.1016/0022-4898(68)90052-9

Bergström, S., 1976. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. Smhi RHO 7, 134. https://doi.org/10.1093/bjsw/bcn100

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834

Bilmes, J.A., 1998. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, International Computer Science Institute. https://doi.org/10.1016/S0550-3213(97)00753-0

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer. https://doi.org/10.1117/1.2819119

Bloomfield, J.P., Marchant, B.P., Bricker, S.H., Morgan, R.B., 2015. Regional analysis of groundwater droughts using hydrograph classification. Hydrol. Earth Syst. Sci. 19, 4327–4344. https://doi.org/10.5194/hess-19-4327-2015

Blunsom, P., 2004. Hidden markov models. Lect. notes 15, 18–19. https://doi.org/10.1007/978-1-61779-400-1_22

Bober, W., 2013. Introduction to Numerical and Analytical Methods with MATLAB for Engineers and Scientists. https://doi.org/https://doi.org/10.1201/b16030

Botter, G., Porporato, A., Rodriguez-Iturbe, I., Rinaldo, A., 2007. Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. Water Resour. Res. 43, 1–14. https://doi.org/10.1029/2006WR005043

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T., 2000. Copulas for Finance - A Reading Guide and Some Applications. SSRN Electron. J. https://doi.org/10.2139/ssrn.1032533

Bray, M., Han, D., 2004. Identification of support vector machines for runoff modelling. J. Hydroinformatics 265–280. https://doi.org/10.2166/hydro.2004.0020

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., Levizzani, V., 2014. Journal of Geophysical Research : Atmospheres rainfall from satellite soil moisture data. J. Geophys. Res. 1–14. https://doi.org/10.1002/2014JD021489

Brocca, L., Crow, W.T., Ciabatta, L., Massari, C., De Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., Wagner, W., 2017. A Review of the Applications of ASCAT Soil Moisture Products. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 2285–2306. https://doi.org/10.1109/JSTARS.2017.2651140

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., Bittelli, M., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. Remote Sens. Environ. 115, 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003

Brocca, L., Melone, F., Moramarco, T., Wager, W., Naeimi, V., Bartalis, Z., Hasenauer, S., 2010a. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol. Earth Syst. Sci. 14, 1881–1893. https://doi.org/10.5194/hess-14-1881-2010

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Hasenauer, S., 2010b. ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy. Remote Sens. Environ. 114, 2745–2755. https://doi.org/10.1016/j.rse.2010.06.009

Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., Hahn, S., 2012. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. IEEE Trans. Geosci. Remote Sens. 50, 2542–2555. https://doi.org/10.1109/TGRS.2011.2177468

Brunetti, G., Šimůnek, J., Bogena, H., Baatz, R., Huisman, J.A., Dahlke, H., Vereecken, H., 2019. On the Information Content of Cosmic-Ray Neutron Data in the Inverse Estimation of Soil Hydraulic Properties. Vadose Zo. J. 18, 0. https://doi.org/10.2136/vzj2018.06.0123

Burnash, R.J., Ferral, R.L., McGuire, R.A., 1973. A Generalized Streamflow Simulation System: Conceptual Models for Digital Computers. Joint Federal State River Forecast Center, Sacramento, CA.

Busch, F.A., Niemann, J.D., Coleman, M., 2012. Evaluation of an empirical orthogonal function-based method to downscale soil moisture patterns based on topographical attributes. Hydrol. Process. 26, 2696–2709. https://doi.org/10.1002/hyp.8363

Butts, M., Graham, D., 2005. Flexible Integrated Watershed Modeling with MIKE SHE, in: Frevert, V.P.S. and D.K. (Ed.), Watershed Models. CRC Press, FL, USA, pp. 245–272.

Carreau, J., Naveau, P., Sauquet, E., 2009. A statistical rainfall-runoff mixture model with heavy-tailed components. Water Resour. Res. 45. https://doi.org/10.1029/2009WR007880

Ceballos, A., Scipal, K., Wagner, W., Martınez-Fernandez, J., 2005. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain. Hydrol. Process. 19, 1549–1566. https://doi.org/10.1002/hyp.5585

Champagne, C., Berg, A., Belanger, J., McNairn, H., de Jeu, R., 2010. Evaluation of soil moisture derived from passive microwave remote sensing over agricultural sites in Canada using ground-based soil moisture monitoring networks. Int. J. Remote Sens. 31, 3669–3690. https://doi.org/10.1080/01431161.2010.483485

Chauhan, N.S., Miller, S., Ardanuy, P., 2003. Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach. Int. J. Remote Sens. 24, 4599–4622. https://doi.org/10.1080/0143116031000156837

Cho, E., Moon, H., Choi, M., 2015. First Assessment of the Advanced Microwave Scanning Radiometer 2 (AMSR2) Soil Moisture Contents in Northeast Asia. J. Meteorol. Soc. Japan. Ser. II 93, 117–129. https://doi.org/10.2151/jmsj.2015-008

Ciabatta, L., Massari, C., Brocca, L., Gruber, A., Reimer, C., Hahn, S., Paulik, C., Dorigo, W., Kidd, R., Wagner, W., 2018. SM2RAIN-CCI: A new global long-term rainfall data set derived from ESA CCI soil moisture. Earth Syst. Sci. Data 10, 267–280.

https://doi.org/10.5194/essd-10-267-2018

Cioffi, F., Conticello, F., Lall, U., Marotta, L., Telesca, V., 2017. Large scale climate and rainfall seasonality in a Mediterranean Area: Insights from a non-homogeneous Markov model applied to the Agro-Pontino plain. Hydrol. Process. 31, 668–686. https://doi.org/10.1002/hyp.11061

Clayton, D.G., 1978. A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. Biometrika 65, 141–151.

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S.B., Cosh, M.H., Dunbar, R.S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J.P., Wu, X., Monerris, A., O'Neill, P.E., Entekhabi, D., Njoku, E.G., Yueh, S., 2017. Validation of SMAP surface soil moisture products with core validation sites. Remote Sens. Environ. 191, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Crow, W.T., Ryu, D., 2008. A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals. Hydrol. Earth Syst. Sci. Discuss. 5, 2005–2044. https://doi.org/10.5194/hessd-5-2005-2008

Curran, J.H., 2012. Streamflow Record Extension for Selected Strams in the Susitan River Basin, Alaska, U.S. Geological Survey Scientific Investigations Report 2012-5210.

Dai, A., 2011. Drought under global warming: A review. Wiley Interdiscip. Rev. Clim. Chang. 2, 45–65. https://doi.org/10.1002/wcc.81

Das, N.N., Entekhabi, D., Njoku, E.G., 2011. An Algorithm for Merging SMAP Radiometer and Radar Data for High Resolution Soil Moisture Retrieval. IEEE Trans. Geosci. Remote Sens. in press, 1–9. https://doi.org/10.1109/TGRS.2010.2089526

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. 39, 1–38. https://doi.org/http://dx.doi.org/10.2307/2984875

Devia, G.K., Ganasri, B.P., Dwarakish, G.S., 2015. A Review on Hydrological Models. Aquat.

Procedia 4, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Dharssi, I., Bovis, K.J., Macpherson, B., Jones, C.P., 2011. Operational assimilation of ASCAT surface soil wetness at the Met Office. Hydrol. Earth Syst. Sci. 15, 2729–2746. https://doi.org/10.5194/hess-15-2729-2011

Djamai, N., Magagi, R., Goïta, K., Merlin, O., Kerr, Y., Roy, A., 2016. A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days. Remote Sens. Environ. 184, 1–14. https://doi.org/10.1016/j.rse.2016.06.010

Djerbouai, S., Souag-Gamane, D., 2016. Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the Algerois Basin in North Algeria. Water Resour. Manag. 30, 2445–2464. https://doi.org/10.1007/s11269-016-1298-6

Dorigo, W., De Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., Fernández-Prieto, D., 2012. Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture. Geophys. Res. Lett. 39, 3–9. https://doi.org/10.1029/2012GL052988

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P.D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S.I., Smolander, T., Lecomte, P., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sens. Environ. 203, 185–215. https://doi.org/10.1016/j.rse.2017.07.001

Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., Kidd, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sens. Environ. 162, 380–395. https://doi.org/10.1016/j.rse.2014.07.023

Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Van Oevelen, P., Robock, A., Jackson, T., 2011. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. Hydrol. Earth Syst. Sci. 15, 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Dorigo, W.A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., Drusch, M., 2013. Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. Vadose Zo. J. 12, 0. https://doi.org/10.2136/vzj2012.0097

Dracup, J.A., Lee, K.S., Paulson, E.G., 1980. On the definition of droughts. Water Resour. Res. 16, 297–302. https://doi.org/10.1029/WR016i002p00297

Drusch, M., Wood, E.F., Gao, H., 2005. Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. Geophys. Res. Lett. 32, 32–35. https://doi.org/10.1029/2005GL023623

Ekanayake, E., Perera, K., 2014. Analysis of Drought Severity and Duration Using Copulas in

Anuradhapura, Sri Lanka. Br. J. Environ. Clim. Chang. 4, 312–327. https://doi.org/10.9734/BJECC/2014/14482

Enenkel, M., Steiner, C., Mistelbauer, T., Dorigo, W., Wagner, W., See, L., Atzberger, C., Schneider, S., Rogenhofer, E., 2016. A combined satellite-derived drought indicator to support humanitarian aid organizations. Remote Sens. 8. https://doi.org/10.3390/rs8040340

Engman, E.T., Gurney, R.J., 1991. Remote sensing in hydrology. Remote Sensing Applications Series. Chapman & Hall, London etc., p. 225.

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98, 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Entekhabi, D., Rodriguez-Iturbe, I., Castelli, F., 1996. Mutual interaction of soil moisture state and atmospheric processes. J. Hydrol. 184, 3–17. https://doi.org/10.1016/0022-1694(95)02965-6

Fang, B., Lakshmi, V., 2014. Soil moisture at watershed scale: Remote sensing techniques. J. Hydrol. 516, 258–272. https://doi.org/10.1016/j.jhydrol.2013.12.008

Farahmand, A., AghaKouchak, A., 2015. A generalized framework for deriving nonparametric standardized drought indicators. Adv. Water Resour. 76, 140–145. https://doi.org/10.1016/j.advwatres.2014.11.012

Favre, A.C., Adlouni, S. El, Perreault, L., Thiémonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. Water Resour. Res. 40, 1–12. https://doi.org/10.1029/2003WR002456

Fernando, A.K., Shamseldin, A.Y., Abrahart, R.J., 2012. Use of Gene Expression Programming for Multimodel Combination of Rainfall-Runoff Models. J. Hydrol. Eng. 17, 975–985. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000533

Ford, T., Labosier, C.F., 2014. Spatial patterns of drought persistence in the Southeastern United States. Int. J. Climatol. 34, 2229–2240. https://doi.org/10.1002/joc.3833

Ford, T.W., Harris, E., Quiring, S.M., 2014. Estimating root zone soil moisture using near-surface observations from SMOS. Hydrol. Earth Syst. Sci. 18, 139–154. https://doi.org/10.5194/hess-18-139-2014

Fumikazu, N., Toshisuke, M., Yoshio, H., Hiroshi, T., Kimihito, N., 2013. Evaluation of water resources by snow storage using water balance and tank model method in the Tedori River basin of Japan. Paddy Water Environ. 11, 113–121. https://doi.org/10.1007/s10333-011-0297-8

Ganguli, P., Ganguly, A.R., 2016. Space-time trends in U.S. meteorological droughts. J. Hydrol. Reg. Stud. 8, 235–259. https://doi.org/10.1016/j.ejrh.2016.09.004

Gauvain, J.-L., Lee, C.-H., 1994. Maximum a posteriori estimation for multivariate Gaussian

mixture observations of Markov chains. IEEE Trans. Speech Audio Process. 2, 291–298. https://doi.org/10.1109/89.279278

Gharhramani, Z., 2001. An Introduction to Hidden Markov Models and Bayesian Networks. J. Pattern Recognit. Artif. Intell. 15, 9–42.

González-Zamora, Á., Sánchez, N., Martínez-Fernández, J., Wagner, W., 2016. Root-zone plant available water estimation using the SMOS-derived soil water index. Adv. Water Resour. 96, 339–353. https://doi.org/10.1016/j.advwatres.2016.08.001

Granata, F., Gargano, R., de Marinis, G., 2016. Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model. Water 8, 69. https://doi.org/10.3390/w8030069

Griesfeller, a., Lahoz, W. a., Jeu, R.A.M. d., Dorigo, W., Haugen, L.E., Svendby, T.M., Wagner, W., 2016. Evaluation of satellite soil moisture products over Norway using ground-based observations. Int. J. Appl. Earth Obs. Geoinf. 45, 155–164. https://doi.org/10.1016/j.jag.2015.04.016

Guerrero-Salazar, P., Jevjevich, V., 1975. Analysis of Drought Characteristics by The Theory of Runs. Hydrol. Pap. 80 1–44.

Güner Bacanli, Ü., 2017. Trend analysis of precipitation and drought in the Aegean region, Turkey. Meteorol. Appl. 24, 239–249. https://doi.org/10.1002/met.1622

Guttman, N.B., 1999. Accepting the Standardized Precipitation Index: a Calculation Algorithm1. JAWRA J. Am. Water Resour. Assoc. 35, 311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

Hain, C.R., Crow, W.T., Mecikalski, J.R., Anderson, M.C., Holmes, T., 2011. An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. J. Geophys. Res. Atmos. 116, 1–18. https://doi.org/10.1029/2011JD015633

Han, E., Heathman, G.C., Merwade, V., Cosh, M.H., 2012a. Application of observation operators for field scale soil moisture averages and variances in agricultural landscapes. J. Hydrol. 444–445, 34–50. https://doi.org/10.1016/j.jhydrol.2012.03.035

Han, E., Merwade, V., Heathman, G.C., 2012b. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. J. Hydrol. 416–417, 98–117. https://doi.org/10.1016/j.jhydrol.2011.11.039

Hao, Z., AghaKouchak, A., 2013. Multivariate Standardized Drought Index: A parametric multi-index model. Adv. Water Resour. 57, 12–18. https://doi.org/10.1016/j.advwatres.2013.03.009

Hao, Z., Singh, V.P., 2015. Drought characterization from a multivariate perspective: A review. J. Hydrol. 527, 668–678. https://doi.org/10.1016/j.jhydrol.2015.05.031

Hayes, M., Svoboda, M., Wall, N., Widhalm, M., 2011. The lincoln declaration on drought indices:

Universal meteorological drought index recommended. Bull. Am. Meteorol. Soc. 92, 485–488. https://doi.org/10.1175/2010BAMS3103.1

Hosseini, S.M., Mahjouri, N., 2016. Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling. Appl. Soft Comput. J. 38, 329–345. https://doi.org/10.1016/j.asoc.2015.09.049

Hughes, J.P., Guttorp, P., 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour. Res. 30, 1535–1546. https://doi.org/10.1029/93WR02983

Hughes, J.P., Guttorp, P., Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. J. R. Stat. Soc. Ser. C (Applied Stat. 48, 15–30. https://doi.org/10.1111/1467-9876.00136

Im, J., Park, S., Rhee, J., Baik, J., Choi, M., 2016. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. Environ. Earth Sci. 75, 1–19. https://doi.org/10.1007/s12665-016-5917-6

Jakeman, A.J., 1993. How Much Complexity Is Warranted in a Rainfall-Runoff Model ? are good predictors of streamflow and. Water Resour. Res. 29, 2637–2649. https://doi.org/10.1029/93WR00877

Jarvis, N.J., Leeds-Harrison, P.B., 1987. Some problems associated with the use of the neutron probe in swelling/shrinkling clay soils. J. Soil Sci. 38, 149–156. https://doi.org/10.1111/j.1365-2389.1987.tb02132.x

JAXA, 2013. GCOM-W1 SHIZUKU Data Users Handbook First Edition.

Joe, H., 1997. NoMultivariate Models and Multivariate Dependence Concepts. Chapman and Hall, New York.

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J.C.J.H., Mechler, R., Botzen, W.J.W., Bouwer, L.M., Pflug, G., Rojas, R., Ward, P.J., 2014. Increasing stress on disaster-risk finance due to large floods. Nat. Clim. Chang. 4, 264–268. https://doi.org/10.1038/nclimate2124

Kalin, L., Isik, S., Schoonover, J.E., Lockaby, B.G., 2010. Predicting Water Quality in Unmonitored Watersheds Using Artificial Neural Networks. J. Environ. Qual. 39, 1429. https://doi.org/10.2134/jeq2009.0441

Kao, S.C., Govindaraju, R.S., 2010. A copula-based joint deficit index for droughts. J. Hydrol. 380, 121–134. https://doi.org/10.1016/j.jhydrol.2009.10.029

Karavitis, C.A., Alexandris, S., Tsesmelis, D.E., Athanasopoulos, G., 2011. Application of the Standardized Precipitation Index (SPI) in Greece. Water (Switzerland) 3, 787–805. https://doi.org/10.3390/w3030787

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. Al, Cabot, F., Gruhier, C., Juglea, S.E., Leroux, D., Mialon, A., Delwart, S., 2012. The SMOS

Soil Moisture Retrieval Algorithm. Geosci. Remote Sens. 50, 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548

Khalil, A.F., Kwon, H.-H., Lall, U., Kaheil, Y.H., 2010. Predictive downscaling based on non-homogeneous hidden Markov models. Hydrol. Sci. J. 55, 333–350. https://doi.org/10.1080/02626661003780342

Kim, S., Liu, Y.Y., Johnson, F.M., Parinussa, R.M., Sharma, A., 2015. A global comparison of alternate AMSR2 soil moisture products: Why do they differ? Remote Sens. Environ. 161, 43–62. https://doi.org/10.1016/j.rse.2015.02.002

Kirshner, S., 2005. Modeling of multivariate time series using hidden Markov models. University of California, Irvine.

Kisi, O., Parmar, K.S., 2016. Application of multivariate adaptive regression spline models in long term prediction of river water pollution. J. Hydrol. 534, 104–112. https://doi.org/10.1016/j.jhydrol.2015.12.014

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.

Kolassa, J., Gentine, P., Prigent, C., Aires, F., 2016. Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy. Part 1: Satellite data analysis. Remote Sens. Environ. 173, 1–14. https://doi.org/10.1016/j.rse.2015.11.011

Kornelsen, K.C., Coulibaly, P., 2015. Reducing multiplicative bias of satellite soil moisture retrievals. Remote Sens. Environ. 165, 109–122. https://doi.org/10.1016/j.rse.2015.04.031

Kumar, R., Musuuza, J.L., Van Loon, A.F., Teuling, A.J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., Attinger, S., 2016. Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator. Hydrol. Earth Syst. Sci. 20, 1117–1131. https://doi.org/10.5194/hess-20-1117-2016

Kwon, H.H., Lall, U., Kim, S.J., 2016. The unusual 2013–2015 drought in South Korea in the context of a multicentury precipitation record: Inferences from a nonstationary, multivariate, Bayesian copula model. Geophys. Res. Lett. 43, 8534–8544. https://doi.org/10.1002/2016GL070270

Kwon, H.H., Lall, U., Obeysekera, J., 2009. Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida. Stoch. Environ. Res. Risk Assess. 23, 879–896. https://doi.org/10.1007/s00477-008-0270-2

Kwon, H.H., Sivakumar, B., Moon, Y. Il, Kim, B.S., 2011. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. Stoch. Environ. Res. Risk Assess. 25, 567–581. https://doi.org/10.1007/s00477-010-0422-z

Lakshmanan, V., Kain, J.S., 2010. A Gaussian Mixture Model Approach to Forecast Verification. Weather Forecast. 25, 908–920. https://doi.org/10.1175/2010WAF2222355.1

Lall, U., Devineni, N., Kaheil, Y., 2016. An Empirical, Nonparametric Simulator for Multivariate

Random Variables with Differing Marginal Densities and Nonlinear Dependence with Hydroclimatic Applications. Risk Anal. 36, 57–73. https://doi.org/10.1111/risa.12432

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241. https://doi.org/10.1029/1998WR900018

Li, C., Singh, V.P., Mishra, A.K., 2013. A bivariate mixed distribution with a heavy-tailed component and its application to single-site daily rainfall simulation. Water Resour. Res. 49, 767–789. https://doi.org/10.1002/wrcr.20063

Li, X., Cheng, H., Tan, L., Ban, F., Sinha, A., Duan, W., Li, H., Zhang, H., Ning, Y., Kathayat, G., Edwards, R.L., 2017. The East Asian summer monsoon variability over the last 145 years inferred from the Shihua Cave record, North China. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-07251-3

Lievens, H., Tomer, S.K., Al Bitar, A., De Lannoy, G.J.M., Drusch, M., Dumedah, G., Hendricks Franssen, H.J., Kerr, Y.H., Martens, B., Pan, M., Roundy, J.K., Vereecken, H., Walker, J.P., Wood, E.F., Verhoest, N.E.C., Pauwels, V.R.N., 2015. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. Remote Sens. Environ. 168, 146–162. https://doi.org/10.1016/j.rse.2015.06.025

Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., De Jeu, R.A.M., Wagner, W., M. Van Dijk, A.I.J., McCabe, M.F., Evans, J.P., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sci. 15, 425–436. https://doi.org/10.5194/hess-15-425-2011

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., Casalí, J., 2018. On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. Adv. Water Resour. 111, 86–104. https://doi.org/10.1016/j.advwatres.2017.10.034

Massari, C., Brocca, L., Barbetta, S., Papathanasiou, C., Mimikou, M., Moramarco, T., 2014. Using globally available soil moisture indicators for flood modelling in Mediterranean catchments. Hydrol. Earth Syst. Sci. 18, 839–853. https://doi.org/10.5194/hess-18-839-2014

Massari, C., Brocca, L., Ciabatta, L., Moramarco, T., Gabellani, S., Albergel, C., De Rosnay, P., Puca, S., Wagner, W., 2015. The Use of H-SAF Soil Moisture Products for Operational Hydrology: Flood Modelling over Italy. Hydrology 2, 2–22. https://doi.org/10.3390/hydrology2010002

Massari, C., Camici, S., Ciabatta, L., Brocca, L., 2018. Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction. Remote Sens. 10. https://doi.org/10.3390/rs10020292

McIntyre, N., Lee, H., Wheater, H., Young, A., Wagener, T., 2005. Ensemble predictions of runoff in ungauged catchments. Water Resour. Res. 41, 1–14. https://doi.org/10.1029/2005WR004289

Mckee, T.B., Doesken, N.J., Kleist, J., 1993. The relationship of drought frequency and duration to time scales. AMS 8th Conf. Appl. Climatol. 179–184. https://doi.org/citeulike-article-id:10490403

Mehrotra, R., Sharma, A., 2005. A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. J. Geophys. Res. D Atmos. 110, 1–13. https://doi.org/10.1029/2004JD005677

Meng, L., Ford, T., Guo, Y., 2017. Logistic regression analysis of drought persistence in East China. Int. J. Climatol. 37, 1444–1455. https://doi.org/10.1002/joc.4789

Merlin, O., Rüdiger, C., Al Bitar, A., Richaume, P., Walker, J.P., Kerr, Y.H., 2012. Disaggregation of SMOS soil moisture in Southeastern Australia. IEEE Trans. Geosci. Remote Sens. 50, 1556–1571. https://doi.org/10.1109/TGRS.2011.2175000

Min, S.K., Kwon, W.T., Park, E.H., Choi, Y., 2003. Spatial and temporal comparisons of droughts over Korea with East Asia. Int. J. Climatol. 23, 223–233. https://doi.org/10.1002/joc.872

Mirabbasi, R., Anagnostou, E.N., Fakheri-Fard, A., Dinpashoh, Y., Eslamian, S., 2013. Analysis of meteorological drought in northwest Iran using the Joint Deficit Index. J. Hydrol. 492, 35–48. https://doi.org/10.1016/j.jhydrol.2013.04.019

Mirabbasi, R., Fakheri-Fard, A., Dinpashoh, Y., 2012. Bivariate drought frequency analysis using the copula method. Theor. Appl. Climatol. 108, 191–206. https://doi.org/10.1007/s00704-011-0524-7

Mo, K.C., 2011. Drought onset and recovery over the United States. J. Geophys. Res. Atmos. 116, 1–14. https://doi.org/10.1029/2011JD016168

Mo, K.C., Schemm, J.E., 2008. Droughts and persistent wet spells over the United States and Mexico. J. Clim. 21, 980–994. https://doi.org/10.1175/2007JCLI1616.1

Mojena, R., 1977. Hierarchical grouping methods and stopping rules: An evaluation. Coputer J. 20, 359–363.

MOLIT, 2016. National Water Resources Long-term Plan.

Moore, R.J., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrol. Sci. J. 30, 273–297. https://doi.org/10.1080/02626668509490989

Moran, M.S., Peters-Lidard, C.D., Watts, J.M., Mc Elroy, S., 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. Can. J. Remote Sens. 30, 805–826. https://doi.org/10.5589/m04-043

Nam, W.H., Hayes, M.J., Svoboda, M.D., Tadesse, T., Wilhite, D.A., 2015. Drought hazard assessment in the context of climate change for South Korea. Agric. Water Manag. 160, 106–117. https://doi.org/10.1016/j.agwat.2015.06.029

Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring.

Agric. For. Meteorol. 133, 69–88. https://doi.org/10.1016/j.agrformet.2005.07.012

Nash, E., Sutcliffe, V., 1970. River flow forecasting through conceptual models Part I - A discussion of principles. J. Hydrol. 10, 282–290.

Neitsch, S.., Arnold, J.., Kiniry, J.., Williams, J.., 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009, Texas Water Resources Institute Technical Report No. 406. Texas.

Nelder, J.A., Mead, R., 1965. A Simplex Method for Function Minimization. Comput. J. 7, 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nelsen, R.B., 1999. An Introduction to Copulas. Springer, New York.

Nguyen, H.H., Jeong, J., Choi, M., 2019. Extension of cosmic-ray neutron probe measurement depth for improving field scale root-zone soil moisture estimation by coupling with representative in-situ sensors. J. Hydrol. 571, 679–696. https://doi.org/10.1016/j.jhydrol.2019.02.018

Okkan, U., Serbes, Z.A., 2012. Rainfall-runoff modeling using least squares support vector machines. Environmetrics 23, 549–564. https://doi.org/10.1002/env.2154

Oliva, R., Daganzo-Eusebio, E., Kerr, Y.H., Mecklenburg, S., Nieto, S., Richaume, P., Gruhier, C., 2012. SMOS radio frequency interference scenario: Status and actions taken to improve the RFI environment in the 1400-1427-MHZ passive band. IEEE Trans. Geosci. Remote Sens. 50, 1427–1439. https://doi.org/10.1109/TGRS.2012.2182775

Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M., 2015. Does model performance improve with complexity? A case study with three hydrological models. J. Hydrol. 523, 147–159. https://doi.org/10.1016/j.jhydrol.2015.01.044

Paik, K., Kim, J.H., Kim, H.S., Lee, D.R., 2005. A conceptual rainfall-runoff model considering seasonal variation. Hydrol. Process. 19, 475–476. https://doi.org/10.1002/hyp.5984

Paiva, A.R. De, Jr, J.H., Orville, R.E., 2016. Journal of Geophysical Research : Atmospheres from high-speed videos 8493–8505. https://doi.org/10.1002/2016JD025137.Received

Paquette, M., Fortier, D., Vincent, W., 2016. Water tracks in the High Arctic: A hydrological network dominated by rapid subsurface flow through patterned ground Journal: Arct. Sci. 353, 334–353.

Park, Seonyoung, Park, Sumin, Im, J., Rhee, J., Shin, J., Park, J., 2017. Downscaling GLDAS Soil Moisture Data in East Asia through Fusion of Multi-Sensors by Optimizing Modified Regression Trees. Water 9, 332. https://doi.org/10.3390/w9050332

Paulik, C., Dorigo, W., Wagner, W., Kidd, R., 2014. Validation of the ASCAT soil water index using in situ data from the International Soil moisture network. Int. J. Appl. Earth Obs. Geoinf. 30, 1–8. https://doi.org/10.1016/j.jag.2014.01.007

Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite

remotely sensed soil moisture. Rev. Geophys. 1–26. https://doi.org/10.1002/2016RG000543

Peng, J., Loew, A., Zhang, S., Wang, J., 2016. Spatial downscaling of global satellite soil moisture data using temperature vegetation dryness index. IEEE Trans. Geosci. Remote Sens. 1, 558–566.

Peng, J., Niesel, J., Loew, A., 2015. Evaluation of soil moisture downscaling using a simple thermal-based proxy-the REMEDHUS network (Spain) example. Hydrol. Earth Syst. Sci. 19, 4765–4782. https://doi.org/10.5194/hess-19-4765-2015

Petropoulos, G.P., Griffiths, H.M., Dorigo, W., Xaver, A., Gruber, A., 2014. Surface Soil Moisture Estimation: Significance, Controls, and Conventional Measurement Techniques. Remote Sens. Energy Fluxes Soil Moisture Content 29–48.

Petropoulos, G.P., Ireland, G., Barrett, B., 2015. Surface soil moisture retrievals from remote sensing: Current status, products & future trends. Phys. Chem. Earth 83–84, 36–56. https://doi.org/10.1016/j.pce.2015.02.009

Pfannerstill, M., Guse, B., Fohrer, N., 2014. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. J. Hydrol. 510, 447–458. https://doi.org/10.1016/j.jhydrol.2013.12.044

Piles, M., Camps, A., Vall-llossera, M., Corbella, I., Panciera, R., Rüdiger, C., Kerr, Y.H., Walker, J., 2011. Downscaling SMOS-Derived Soil Moisture Using MODIS Visible / Infrared Data. IEEE Trans. Geosci. Remote Sens. 49, 3156–3166.

Piles, M., Sánchez, N., Vall-Llossera, M., Camps, A., Martínez-Fernandez, J., Martinez, J., Gonzalez-Gambau, V., 2014. A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian peninsula. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 3845–3857. https://doi.org/10.1109/JSTARS.2014.2325398

Qiu, J., Crow, W.T., Nearing, G.S., Mo, X., Liu, S., 2014. The impact of vertical measurement depth on the information content of soil moisture times series data. Geophys. Res. Lett. 41, 4997–5004. https://doi.org/10.1002/2014GL060017

Rabiner, L.R., 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE. https://doi.org/10.1109/5.18626

Raghavendra, S., Deka, P.C., 2014. Support vector machine applications in the field of hydrology: A review. Appl. Soft Comput. J. 19, 372–386. https://doi.org/10.1016/j.asoc.2014.02.002

Rahmani, A., Golian, S., Brocca, L., 2016. Multiyear monitoring of soil moisture over Iran through satellite and reanalysis soil moisture products. Int. J. Appl. Earth Obs. Geoinf. 48, 85–95. https://doi.org/10.1016/j.jag.2015.06.009

Ranney, K.J., Niemann, J.D., Lehman, B.M., Green, T.R., Jones, A.S., 2015. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. Adv. Water Resour. 76, 81–96. https://doi.org/10.1016/j.advwatres.2014.12.003

Reichle, R.H., Koster, R.D., Liu, P., Mahanama, S.P.P., Njoku, E.G., Owe, M., 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). J. Geophys. Res. Atmos. 112, 1–14. https://doi.org/10.1029/2006JD008033

Renard, B., Lang, M., 2007. Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. Adv. Water Resour. 30, 897–912. https://doi.org/10.1016/j.advwatres.2006.08.001

Renzullo, L.J., van Dijk, A.I.J.M., Perraud, J.M., Collins, D., Henderson, B., Jin, H., Smith, A.B., McJannet, D.L., 2014. Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. J. Hydrol. 519, 2747–2762. https://doi.org/10.1016/j.jhydrol.2014.08.008

Requena, A.I., Mediero, L., Garrote, L., 2013. A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. Hydrol. Earth Syst. Sci. 17, 3023–3038. https://doi.org/10.5194/hess-17-3023-2013

Ridolfi, L., D'Odorico, P., Porporato, A., Rodriguez-Iturbe, I., 2003. Stochastic soil moisture dynamics along a hillslope. J. Hydrol. 272, 264–275. https://doi.org/10.1016/S0022-1694(02)00270-6

Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011607

Robertson, A.W., Kirshner, S., Smyth, P., 2004. Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. J. Clim. 17, 4407–4424. https://doi.org/10.1175/JCLI-3216.1

Robertson, A.W., Kirshner, S., Smyth, P., Charles, S.P., Bates, B.C., 2006. Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland. Q. J. R. Meteorol. Soc. 132, 519–542. https://doi.org/10.1256/qj.05.75

Robertson, A.W., Sergey, K., Padhraic, S., 2003. Hidden Markov models for modeling daily rainfall occurrence over Brazil, Technical Report UCI-ICS 03-27. Information and Computer Science University of Califonia, Irvine.

Rui, H., Beaudoing, H.K., 2017. README Document for GLDAS Version 2 Data Products 1–32.

Rüschendorf, L., 2009. On the distributional transform, Sklar's theorem, and the empirical copula process. J. Stat. Plan. Inference 139, 3921–3927. https://doi.org/10.1016/j.jspi.2009.05.030

Ryu, D., Famiglietti, J.S., 2005. Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment. Water Resour. Res. 41, 1–13. https://doi.org/10.1029/2004WR003835

S.U, S.L., Singh, D.N., Shojaei, M., 2014. A critical review of soil moisture measurement.

Measurement 54, 92–105. https://doi.org/10.1016/j.measurement.2014.04.007

Sabater, J.M., Jarlan, L., Calvet, J.-C., Bouyssel, F., De Rosnay, P., 2007. From Near-Surface to Root-Zone Soil Moisture Using Different Assimilation Techniques. J. Hydrometeorol. 8, 194–206. https://doi.org/10.1175/JHM571.1

Sadeghi, A.M., Hancock, G.D., Waite, W.P., Scott, H.D., Rand, J.A., 1984. Microwave Measurements of Moisture Distributions in the Upper Soil Profile. Water Resour. Res. 20, 927–934.

Salvadori, G., De Michele, C., 2004. Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. Water Resour. Res. 40, 1–17. https://doi.org/10.1029/2004WR003133

Samui, P., Kothari, D.P., 2011. Utilization of a least square support vector machine (LSSVM) for slope stability analysis. Sci. Iran. 18, 53–58. https://doi.org/10.1016/j.scient.2011.03.007

Santos, J.F., Pulido-Calvo, I., Portela, M.M., 2010. Spatial and temporal variability of droughts in Portugal. Water Resour. Res. 46, 1–13. https://doi.org/10.1029/2009WR008071

Schmugge, T., 1978. Remote Sensing of Surface Soil Moisture. J. Appl. Meteorol. https://doi.org/10.1175/1520-0450(1978)017<1549:rsossm>2.0.co;2

Schneider, S., Wang, Y., Wagner, W., Mahfouf, J.-F., 2014. Impact of ASCAT Soil Moisture Assimilation on Regional Precipitation Forecasts: A Case Study for Austria. Mon. Weather Rev. 142, 1525–1541. https://doi.org/10.1175/MWR-D-12-00311.1

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.

Scipal, K., Drusch, M., Wagner, W., 2008. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. Adv. Water Resour. 31, 1101–1112. https://doi.org/10.1016/j.advwatres.2008.04.013

Shamshirband, S., Gocić, M., Petković, D., Javidnia, H., Ab Hamid, S.H., Mansor, Z., Qasem, S.N., 2015. Clustering project management for drought regions determination: A case study in Serbia. Agric. For. Meteorol. 200, 57–65. https://doi.org/10.1016/j.agrformet.2014.09.020

Shiau, J.T., Lin, J.W., 2016. Clustering Quantile Regression-Based Drought Trends in Taiwan. Water Resour. Manag. 30, 1053–1069. https://doi.org/10.1007/s11269-015-1210-9

Shrestha, D.L., 2009. Uncertainty Analysis in Rainfall-Runoff Modelling: Application of Machine Learning Techniques, Delft University of Technology and \rof the Academic Board of UNESCO-IHE Institute for Water \rEducation.

Shukla, S., Steinemann, A.C., Lettenmaier, D.P., 2011. Drought Monitoring for Washington State: Indicators and Applications. J. Hydrometeorol. 12, 66–83. https://doi.org/10.1175/2010JHM1307.1

Sittner, W.T., 1976. WMO project on intercomparison of conceptual models used in hydrological forecasting. Hydrol. Sci. Bull. 21, 203–213. https://doi.org/10.1080/02626667609491617

Sivapragasam, C., Liong, S., Pasha, M., 2001. Rainfall and runoff forecasting with SSA-SVM approach. J. Hydroinformatics 3, 141–152. https://doi.org/10.1016/j.jhydrol.2010.02.030

Sklar, M., 1959. Fonctions de Répartition À N Dimensions Et Leurs Marges. Publications de l'Institut de Statistique de L'Université de Paris 8, pp. 229–231.

Smyth, P., Heckerman, D., Jordan, M.I., 1997. Probabilistic Independence Networks for Hidden Markov Probability Models. Neural Comput. 9, 227–269. https://doi.org/10.1162/neco.1997.9.2.227

Song, J.H., Her, Y., Park, J., Lee, K. Do, Kang, M.S., 2017. Simulink Implementation of a Hydrologic Model: A Tank Model Case Study. Water 9, 1–17. https://doi.org/10.3390/w9090639

Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., Vogt, J., 2014. World drought frequency, duration, and severity for 1951-2010. Int. J. Climatol. 34, 2792–2804. https://doi.org/10.1002/joc.3875

Sprenger, M., Seeger, S., Blume, T., Weiler, M., 2016. A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. Water Resour. Res. 52, 5727–5754. https://doi.org/10.1002/2014WR015716

Srivastava, P.K., Han, D., Ramirez, M.R., Islam, T., 2013. Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. Water Resour. Manag. 27, 3127–3144. https://doi.org/10.1007/s11269-013-0337-9

Su, C.H., Ryu, D., Young, R.I., Western, A.W., Wagner, W., 2013. Inter-comparison of microwave satellite soil moisture retrievals over the Murrumbidgee Basin, southeast Australia. Remote Sens. Environ. 134, 1–11. https://doi.org/10.1016/j.rse.2013.02.016

Subash, N., Ram Mohan, H.S., 2011. Trend detection in rainfall and evaluation of standardized precipitation index as a drought assessment index for rice-wheat productivity over IGR in India. Int. J. Climatol. 31, 1694–1709. https://doi.org/10.1002/joc.2188

Sugawara, M., 1979. Automatic calibration of the tank model. Hydrol. Sci. Bull. 24, 375–388. https://doi.org/10.1080/02626667909491876

Suykens, J.A.K., De Brabanter, J., Lukas, L., Vandewalle, J., 2002. Weighted least squares support vector machines: Robustness and sparce approximation. Neurocomputing 48, 85–105. https://doi.org/10.1016/S0925-2312(01)00644-0

Svoboda, M.D., Fuchs, B.A., 2017. Handbook of drought indicators and indices, Drought and Water Crises: Integrating Science, Management, and Policy, Second Edition. World Meteorological Organization (WMO) and Global Water Partnership (GWP), Geneva. https://doi.org/10.1201/b22009

Tatli, H., 2015. Detecting persistence of meteorological drought via the Hurst exponent. Meteorol. Appl. 22, 763–769. https://doi.org/10.1002/met.1519

Topp, G.C., Davis, J.L., Annan, A.P., 1980. Electromagnetic Determination of Soil Water Content: Measruements in Coaxial Transmission Lines. Water Resour. Res. 16, 574–582. https://doi.org/10.1029/WR016i003p00574

Van de Griend, A.A., Engman, E.T., 1985. Partial area hydrology and remote sensing. J. Hydrol. 81, 211–251. https://doi.org/10.1016/0022-1694(85)90033-2

Van Loon, A.F., 2015. Hydrological drought explained. Wiley Interdiscip. Rev. Water 2, 359–392. https://doi.org/10.1002/wat2.1085

Van Loon, A.F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A.I.J.M., Tallaksen, L.M., Hannaford, J., Uijlenhoet, R., Teuling, A.J., Hannah, D.M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., Van Lanen, H.A.J., 2016. Drought in a human-modified world: Reframing drought definitions, understanding, and analysis approaches. Hydrol. Earth Syst. Sci. 20, 3631–3650. https://doi.org/10.5194/hess-20-3631-2016

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vaze, J., Jordan, P., Beecham, R., Frost, A., Summerell, G., 2011. Guidelines for rainfall-runoff modelling: Towards best practice model application, eWater Cooperative Research Center.

Verhoest, N.E.C., Van Den Berg, M.J., Martens, B., Lievens, H., Wood, E.F., Pan, M., Kerr, Y.H., Al Bitar, A., Tomer, S.K., Drusch, M., Vernieuwe, H., De Baets, B., Walker, J.P., Dumedah, G., Pauwels, V.R.N., 2015. Copula-based downscaling of coarse-scale soil moisture observations with implicit bias correction. IEEE Trans. Geosci. Remote Sens. 53, 3507–3521. https://doi.org/10.1109/TGRS.2014.2378913

Vidal, J.-P., Martin, E., Franchistéguy, L., Habets, F., Soubeyroux, J.-M., Blanchard, M., Baillon, M., 2009. Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite. Hydrol. Earth Syst. Sci. Discuss. 6, 6455–6501. https://doi.org/10.5194/hessd-6-6455-2009

Vilasa, L., Miralles, D.G., de Jeu, R.A.M., Dolman, A.J., 2017. Global soil moisture bimodality in satellite observations and climate models. J. Geophys. Res. Atmos. 122, 4299–4311. https://doi.org/10.1002/2016JD026099

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Inf. Theory 13, 260–269. https://doi.org/10.1109/TIT.1967.1054010

Wagner, W., Gruber, A., n.d. Information Content of ASCAT Soil Moisture Data Mission Goal of SMOS and SMAP.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., De Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., Zeil, P., Rubel, F., 2013. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorol. Zeitschrift 22, 5–33. https://doi.org/10.1127/0941-2948/2013/0399

Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS Scatterometer and soil data. Remote Sens. Environ. 70, 191–207. https://doi.org/10.1016/S0034-4257(99)00036-X

Wang, L., Qu, J.J., 2009. Satellite remote sensing applications for surface soil moisture monitoring: A review. Front. Earth Sci. China 3, 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wang, S., Mo, X., Liu, S., Lin, Z., Hu, S., 2016. Validation and trend analysis of ECV soil moisture data on cropland in North China Plain during 1981–2010. Int. J. Appl. Earth Obs. Geoinf. 48, 110–121. https://doi.org/10.1016/j.jag.2015.10.010

WMO, 2012. Standardized Precipitation Index User Guide (WMO-No. 1090) 24. https://doi.org/ISBN 978-92-63-11091-6

Wu, M.C., Lin, G.F., Lin, H.Y., 2014. Improving the forecasts of extreme streamflow by support vector regression with the data extracted by self-organizing map. Hydrol. Process. 28, 386–397. https://doi.org/10.1002/hyp.9584

Wu, Q., Liu, H., Wang, L., Deng, C., 2016. Evaluation of AMSR2 soil moisture products over the contiguous United States using in situ data from the International Soil Moisture Network. Int. J. Appl. Earth Obs. Geoinf. 45, 187–199. https://doi.org/10.1016/j.jag.2015.10.011

Xing, C., Chen, N., Zhang, X., Gong, J., 2017. A Machine Learning Based Reconstruction Method for Satellite Remote Sensing of Soil Moisture Images with In Situ Observations. Remote Sens. 9, 484. https://doi.org/10.3390/rs9050484

Yan, X., Chowdhury, N.A., 2013. Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. Int. J. Electr. Power Energy Syst. 53, 20–26. https://doi.org/10.1016/j.ijepes.2013.04.006

Yang, P., Xia, J., Zhang, Y., Han, J., Wu, X., 2017. Quantile regression and clustering analysis of standardized precipitation index in the Tarim River Basin, Xinjiang, China. Theor. Appl. Climatol. 134, 901–912. https://doi.org/10.1007/s00704-017-2313-4

Yevjevich, V., 1967. An objective approach to definitions and investigations of continental hydrologic droughts. Hydrol. Pap. 23 1–18. https://doi.org/10.1016/0022-1694(69)90110-3

Yoo, J., Kwon, H.H., Kim, T.W., Ahn, J.H., 2012. Drought frequency analysis using cluster analysis and bivariate probability distribution. J. Hydrol. 420–421, 102–111. https://doi.org/10.1016/j.jhydrol.2011.11.046

Yoo, J., Kwon, H.H., So, B.J., Rajagopalan, B., Kim, T.W., 2015. Identifying the role of typhoons as drought busters in South Korea based on hidden Markov chain models. Geophys. Res. Lett. 42, 2797–2804. https://doi.org/10.1002/2015GL063753

Yoo, J.H., 2009. Maximization of hydropower generation through the application of a linear programming model. J. Hydrol. 376, 182–187. https://doi.org/10.1016/j.jhydrol.2009.07.026

Young, C.C., Liu, W.C., Wu, M.C., 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Appl. Soft

Comput. J. 53, 205–216. https://doi.org/10.1016/j.asoc.2016.12.052

Yu, P.S., Chen, S.T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. J. Hydrol. 328, 704–716. https://doi.org/10.1016/j.jhydrol.2006.01.021

Zaman, B., Mckee, M., 2014. Spatio-Temporal Prediction of Root Zone Soil Moisture Using Multivariate Relevance Vector Machines. Open J. Mod. Hydrol. 4, 80–90. https://doi.org/dx.doi.org/10.4236/ojmh.2014.43007

Zargar, A., Sadiq, R., Naser, B., Khan, F.I., 2011. A review of drought indices. Environ. Rev. 19, 333–349. https://doi.org/10.1139/a11-013

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., Zou, P., 2015. Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. Remote Sens. Environ. 163, 91–110. https://doi.org/10.1016/j.rse.2015.03.008

Zhang, L., Singh, V.P., 2006. Bivariate Flood Frequency Analysis Using the Copula Method. Water 11, 150–164.

Zhang, L., Zhou, T., 2015. Drought over East Asia: A review. J. Clim. 28, 3375–3399. https://doi.org/10.1175/JCLI-D-14-00259.1

Zhao, W., Li, A., 2013. A downscaling method for improving the spatial resolution of AMSR-E derived soil moisture product based on MSG-SEVIRI data. Remote Sens. 5, 6790–6811. https://doi.org/10.3390/rs5126790

Zhuo, L., Han, D., 2016. Could operational hydrological models be made compatible with satellite soil moisture observations? Hydrol. Process. 30, 1637–1648. https://doi.org/10.1002/hyp.10804